Dynamic sparsity on dynamic regression models[2]

Paloma V. Schwartzman[1] Hedibert F. Lopes[2]

[1]Data Scientist, iFood, São Paulo, Brazil

[2]Professor of Statistics and Econometrics - INSPER, São Paulo, Brazil

School of Mathematical and Statistical Sciences
Arizona State University
February 2020

[2]Paper and slides can be found in my page at `hedibert.org`

Outline

Outline

Motivation 1: Dynamic linear regression

Consider the (univariate) normal *dynamic linear model* (NDLM) expressed by

$$y_t = x_t'\beta_t + \nu_t, \qquad \nu_t \sim N(0, V_t) \tag{1}$$

$$\beta_t = G_t\beta_{t-1} + \omega_t, \qquad \omega_t \sim N(0, W_t), \tag{2}$$

where β_t is q-dimensional.

- Static regression model: $G_t = I_q$ and $W_t = 0$ for all t.
- Dynamic regression model: $G_t = I_q$ for all t.
  Frühwirth-Schnatter and Wagner [2010], Chan et al. [2012]

Multivariate: $\dim(y_t) = m$

- TVP-VAR(k) model: $dim(y_t) = m$ and $q = dim(x_t) = mk$.
  Koop and Korobilis [2013], Belmonte, Koop, and Korobilis [2014]
- Dynamic factor model: $dim(y_t) = m$, β_t factors and x_t loadings.
  Lopes and Carvalho [2007], Lopes, Salazar, and Gamerman [2008b]

Motivation 2: Time-varying Cholesky decomposition

Let

$$y_t = (y_{1t}, \dots, y_{mt})' \sim N(0, \Sigma_t).$$

Then

$$\Sigma_t = L_t D_t L_t' \quad \text{and} \quad \Sigma_t^{-1} = T_t' D_t^{-1} T_t,$$

where $T_t = L_t^{-1}$ is a unit lower triangular matrix and $D_t = \text{diag}(\sigma_{1t}^2, \dots, \sigma_{mt}^2)$.

It is easy to see that $T_t y_t = \varepsilon_t \sim N(0, D_t)$.

By assuming that the entries of T_t are $-\beta_{ijt}$, it follows that $y_{1t} \sim N(0, \sigma_{1t}^2)$, and

$$y_{it} = \beta_{i1t} y_{1t} + \beta_{i2t} y_{2t} + \cdots + \beta_{i,i-1,t} y_{i-1,t} + \varepsilon_{it} \qquad (q_i = i-1),$$

where $\varepsilon_{it} \sim N(0, \sigma_{it}^2)$ and $i = 2, \dots, m$.

Lopes, McCulloch, and Tsay [2008a] Cholesky stochastic volatility model.
Carvalho, Lopes, and McCulloch [2018] Long run volatility of stocks.

A few references

- Lopes, McCulloch, and Tsay [2008a]
- Frühwirth-Schnatter and Wagner [2010]
- Chan, Koop, Leon-Gonzalez, and Strachan [2012]*
- Nakajima and West [2013]
- Belmonte, Koop, and Korobilis [2014]*
- Kalli and Griffin [2014]*
- Bitto and Frühwirth-Schnatter [2016]*
- Rocková and McAlinn [2018]*
- Kowal, Matteson, and Ruppert [2018]

* Forecasting US inflation rate.

Outline

Ridge and lasso regressions

Let us consider the static regression,

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_q x_{qt} + \nu_t,$$

and RSS$= (y - X\beta)'(y - X\beta)$.

- *Ridge regression* Hoerl and Kennard [1970] - ℓ_2 penalty on β:

  $$\hat{\beta}_{ridge} = \arg\min_{\beta}\{RSS + \lambda_r^2 \sum_{j=1}^{q} \beta_j^2\}, \qquad \lambda_r^2 \geq 0,$$

  leading to $\hat{\beta}_{ridge} = (X'X + \lambda_r^2 I_q)^{-1} X'y$.

- *Lasso regression* Tibshirani [1996] - ℓ_1 penalty on β:

  $$\hat{\beta}_{lasso} = \arg\min_{\beta}\{RSS + \lambda_l \sum_{j=1}^{q} |\beta_j|\}, \qquad \lambda \geq 0,$$

  which can be solved by a *coordinate gradient descent* algorithm.

Ridge and lasso estimates are posterior modes!

The posterior mode or the maximum a posteriori (MAP) is given by

$$\tilde{\beta}_{\text{mode}} = \arg\min_{\beta}\{-2\log p(y|\beta) - 2\log p(\beta)\}$$

The $\hat{\beta}_{ridge}$ estimate equals the posterior mode of the normal linear model with

$$p(\beta_j) \propto \exp\{-0.5\lambda_r^2\beta_j^2\},$$

which is a Gaussian distribution with location 0 and scale $1/\lambda_r^2$, $N(0, 1/\lambda_r^2)$.
The mean is 0, the variance is $1/\lambda_r^2$ and the excess kurtosis is 0.

The $\hat{\beta}_{lasso}$ estimate equals the posterior mode of the normal linear model with

$$p(\beta_j) \propto \exp\{-0.5\lambda_l|\beta_j|\},$$

which is a Laplace distribution with location 0 and scale $2/\lambda_l$, Laplace$(0, 2/\lambda_l)$.
The mean is 0, the variance is $8/\lambda_l^2$ and excess kurtosis is 3.

Spike and slab model (or scale mixture of normals)

Ishwaran and Rao [2005] define a spike and slab model as a Bayesian model specified by the following prior hierarchy:

$$\begin{aligned}(y_t|x_t, \beta, \sigma^2) &\sim N(x_t'\beta, \sigma^2), \qquad t = 1, \ldots, n \\ (\beta|\psi) &\sim N(0, \text{diag}(\psi)) \\ \psi &\sim \pi(d\psi) \\ \sigma^2 &\sim \mu(d\sigma^2)\end{aligned}$$

They go to say that

> *"Lempers [1988] and Mitchell and Beauchamp [1988] were among the earliest to pioneer the spike and slab method. The expression 'spike and slab' referred to the prior for β used in their hierarchical formulation."*

Spike and slab model (or scale mixture of normals model)

Regularization and variable selection are done by assuming independent prior distributions from the SMN class to each coefficient β_j:

$$\beta_j|\psi_j \sim N(0, \psi_j) \qquad \text{and} \qquad \psi_j \sim p(\psi_j)$$

so

$$p(\beta_j) = \int p(\beta_j|\psi_j)p(\psi_j)d\psi_j.$$

Mixing density $p(\psi_j)$	Marginal density $p(\beta_j)$	$V(\beta_j)$	Ex.kurtosis(β_j)
$\psi_j = 1/\lambda_r^2$	$N(0, 1/\lambda_r^2)$ - (ridge)	$1/\lambda_r^2$	0
$IG(\eta/2, \eta\tau^2/2)$	$t_\eta(0, \tau^2)$	$\eta/(\eta-2)\tau^2$	$6/(\eta-4)$
$G(1, \lambda_l^2/8)$	Laplace$(0, 2/\lambda_l)$ - (blasso)	$8/\lambda_l^2$	3
$G(\zeta, 1/(2\gamma^2))$	$NG(\zeta, \gamma^2)$	$2\zeta\gamma^2$	$3/\zeta$

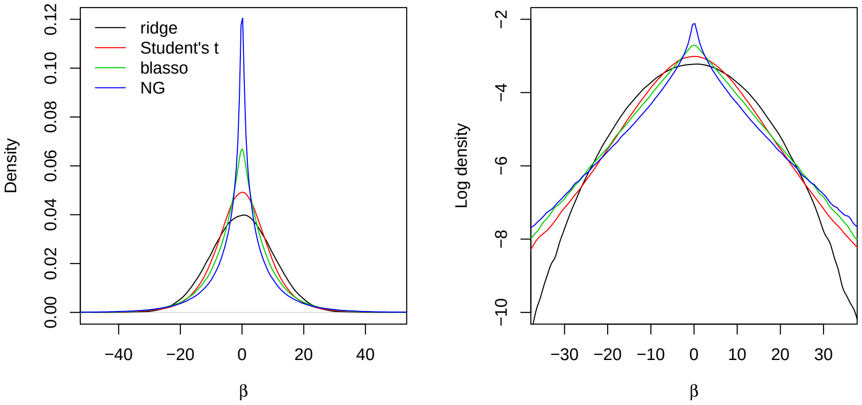
Griffin and Brown [2010] Normal-Gamma prior:

$$p(\beta|\zeta, \gamma^2) = \frac{1}{\sqrt{\pi}2^{\zeta-1/2}\gamma^{\zeta+1/2}\Gamma(\zeta)}|\beta|^{\zeta-1/2}K_{\zeta-1/2}(|\beta|/\gamma),$$

where K is the modified Bessel function of the 3rd kind.

Illustration

Ridge: $\lambda_r^2 = 0.01$ $\Rightarrow$ Excess kurtosis=0
Student's t: $\eta = 5$, $\tau^2 = 60$ $\Rightarrow$ Excess kurtosis=6
Blasso: $\lambda_l^2 = 0.08$ $\Rightarrow$ Excess kurtosis=3
NG: $\xi = 0.5$, $\gamma^2 = 100$ $\Rightarrow$ Excess kurtosis=6
All variances are equal to 100.



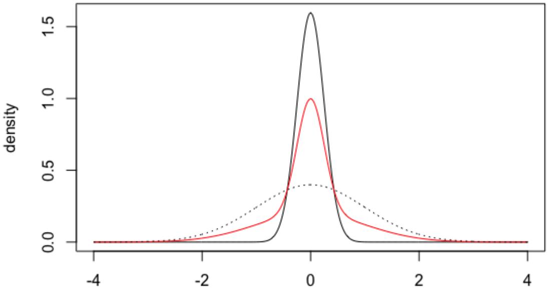
Stochastic search variable selection (SSVS) prior

SSVS George and McCulloch [1993]: For small $\tau > 0$ and $c >> 1$,

$$\beta|\omega,\tau^2,c^2 \sim (1-\omega)\underbrace{N(0,\tau^2)}_{spike} + \omega\underbrace{N(0,c^2\tau^2)}_{slab}.$$

SMN representation: $\beta|\psi \sim N(0,\psi)$ and

$$\psi|\omega,\tau^2,c^2 \sim (1-\omega)\delta_{\tau^2}(\psi) + \omega\delta_{c^2\tau^2}(\psi)$$

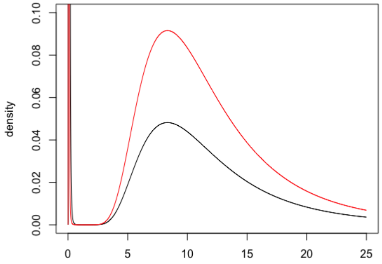


Scaled SSVS prior = normal mixture of IG prior

NMIG prior of Ishwaran and Rao [2005]: For $\upsilon_0 \ll \upsilon_1$,

$$
\begin{aligned}
\beta | K, \tau^2 &\sim N(0, K\tau^2), \\
K | \omega, \upsilon_0, \upsilon_1 &\sim (1-\omega)\delta_{\upsilon_0}(K) + \omega\delta_{\upsilon_1}(K), \\
\tau^2 &\sim IG(a_\tau, b_\tau).
\end{aligned} \tag{3}
$$

- Large ω implies non-negligible effects.
- The scale $\psi = K\tau^2 \sim (1-\omega)IG(a_\tau, \upsilon_0 b_\tau) + \omega IG(a_\tau, \upsilon_1 b_\tau)$.
- $p(\beta)$ is a two component mixture of scaled Student's t distributions.



Other mixture priors

Frühwirth-Schnatter and Wagner [2011]: absolutely continuous priors

$$\beta \sim (1-\omega) p_{spike}(\beta) + \omega p_{slab}(\beta), \tag{4}$$

Let $Q > 0$ a scale parameter and

$$r = \frac{\mathrm{Var}_{spike}(\beta)}{\mathrm{Var}_{slab}(\beta)} \ll 1,$$

then the mixing densities for ψ,

1. IG: $\psi \sim (1-\omega) IG(\nu, rQ) + \omega IG(\nu, Q)$,
2. Exp: $\psi \sim (1-\omega) Exp(1/2rQ) + \omega Exp(1/2Q)$,
3. Gamma: $\psi \sim (1-\omega) G(a, 1/2rQ) + \omega G(a, 1/2Q)$,

leads to the marginal densities for β,

1. Scaled-t: $\beta \sim (1-\omega) t_{2\nu}(0, rQ/\nu) + \omega t_{2\nu}(0, Q/\nu)$,
2. Laplace: $\beta \sim (1-\omega) Lap(\sqrt{rQ}) + \omega Lap(\sqrt{Q})$,
3. NG: $\beta \sim (1-\omega) NG(a, r, Q) + \omega NG(a, Q)$.

Inverted-Gamma prior for the variance of β

It is easy to see that, for a constant c,

$$\mathrm{Var}_{spike}(\beta) = cQr \quad \text{and} \quad \mathrm{Var}_{slab}(\beta) = cQ.$$

Therefore, when

$$v_\beta = \mathrm{Var}(\beta) = (1-\omega)\,\mathrm{Var}_{spike}(\beta) + \omega\,\mathrm{Var}_{slab}(\beta) \sim IG(c_0, C_0),$$

the implied distribution of Q is

$$Q \sim IG\left(c_0, \frac{C_0}{c((1-\omega)r+\omega)}\right).$$

Spike-and-slab priors:

Prior	Spike	Slab	$p(\beta)$	Constant c
SSVS	$\psi = rQ$	$\psi = Q$	$(1-\omega)N(0, rQ) + \omega N(0, Q)$	1
NMIG	$IG(\nu, rQ)$	$IG(\nu, Q)$	$(1-\omega)t_{2\nu}(0, rQ/\nu) + \omega t_{2\nu}(0, Q/\nu)$	$1/(\nu-1)$
Laplaces	$Exp(1/2rQ)$	$Exp(1/2Q)$	$(1-\omega)Lap(\sqrt{rQ}) + \omega Lap(\sqrt{Q})$	2
Normal-Gammas	$G(a, 1/2rQ)$	$G(a, 1/2Q)$	$(1-\omega)NG(\beta_j\|a, r, Q) + \omega NG(\beta_j\|a, Q)$	$2a$
Laplace-t	$Exp(1/2rQ)$	$IG(\nu, Q)$	$(1-\omega)Lap(\sqrt{rQ}) + \omega t_{2\nu}(0, Q/\nu)$	$c_1 = 2, c_2 = 1/(\nu-1)$

Outline

Sparsity in dynamic regressions

Recall,

$$y_t = \beta_{1t}x_{1t} + \beta_{2t}x_{2t} + \cdots + \beta_{qt}x_{qt} + \nu_t,$$

for large q, say 100, 500 or 1000.

- Two main obstacles:
  1. Time-varying parameters (states), and
  2. A large number of predictors q.
- Two sources of sparsity:
  1. Horizontal/static sparsity: $\beta_{jt} = 0, \forall t$ for some coefficients j.
  2. Vertical/dynamic sparsity: $\beta_{jt} = 0$ for several js at time t.

Illustration: $q = 5$ and $T = 12$

	jan	*feb*	*mar*	*apr*	*may*	*jun*	*jul*	*aug*	*sep*	*oct*	*nov*	*dec*
x_1	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{1,6}$	$\beta_{1,7}$	$\beta_{1,8}$	$\beta_{1,9}$	$\beta_{1,10}$	$\beta_{1,11}$	$\beta_{1,12}$
x_2	0	0	0	0	0	0	0	0	0	0	0	0
x_3	$\beta_{3,1}$	$\beta_{3,2}$	$\beta_{3,3}$	$\beta_{3,4}$	$\beta_{3,5}$	0	0	0	$\beta_{3,9}$	$\beta_{3,10}$	$\beta_{3,11}$	$\beta_{3,12}$
x_4	0	0	$\beta_{4,3}$	$\beta_{4,4}$	$\beta_{4,5}$	$\beta_{4,6}$	$\beta_{4,7}$	$\beta_{4,8}$	$\beta_{4,9}$	$\beta_{4,10}$	$\beta_{4,11}$	0
x_5	$\beta_{5,1}$	$\beta_{5,2}$	$\beta_{5,3}$	$\beta_{5,4}$	$\beta_{5,5}$	0	0	0	$\beta_{5,9}$	$\beta_{5,10}$	$\beta_{5,11}$	$\beta_{5,12}$

Horizontal sparsity

Belmonte, Koop, and Korobilis [2014] used a non-centered parametrization:

$$
\begin{aligned}
y_t &= x_t'\beta + \sum_{j=1}^{q} \psi_j^{1/2} \tilde{\beta}_{jt} x_{jt} + \nu_t \\
\tilde{\beta}_{jt} &= \tilde{\beta}_{j,t-1} + u_{jt},
\end{aligned}
\tag{5}
$$

where $\tilde{\beta}_{jt} = \psi_j^{-1/2}\beta_{jt}$ and $u_{jt} \sim N(0,1)$ for $j = 1, \ldots, q$.

- β_j: Laplace prior Park and Casella [2008]

$$\beta_j|\tau_j^2 \sim N(0,\tau_j^2), \qquad \tau_j^2 \sim Exp(\lambda^2/2).$$

- $\beta_{j1}, \ldots, \beta_{jT}$: Laplace prior on standard deviations $\psi_j^{1/2}$

$$\psi_j^{1/2}|\xi_j^2 \sim N(0,\xi_j^2), \qquad \xi_j^2 \sim Exp(\kappa^2/2).$$

Bitto and Frühwirth-Schnatter [2016] adopt a similar strategy by using the Normal-Gamma prior to shrink both β_j and $\psi_j^{1/2}$.

Vertical sparsity

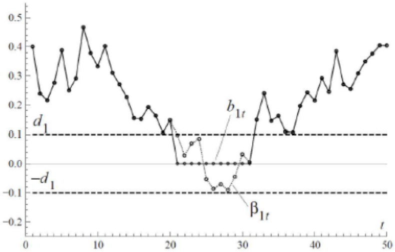
Nakajima and West [2013] - Latent threshold DLMs

For $t = 1, \ldots, T$,

$$y_t = x_{1t} b_{1t} + \ldots + x_{kt} b_{kt} + \varepsilon_t,$$

where, for $i = 1, \ldots, k$, $b_{it} = \beta_{it} s_{it}$, $s_{it} = I(|\beta_{it}| \geq d_i)$, and

$$\beta_{it} | \beta_{i,t-1}, \mu_i, \varphi_i, \psi_i \sim N(\mu_i + \varphi_i(\beta_{it-1} - \mu_i), (1 - \varphi_i^2)\psi_i).$$



Key sparsity prior parameters: $Pr(b_{it} = 0)$

$$d_i | \mu_i, \varphi_i, \psi_i \sim U(0, |\mu_i| + K\psi_i)$$

Vertical sparsity

Kalli and Griffin [2014] - NGAR process

For $t = 1, \ldots, T$,

$$y_t = x_{1t}\beta_{1t} + \ldots + x_{kt}\beta_{kt} + \varepsilon_t,$$

the normal-gamma autoregressive process models β_{jt} as follows:

$$\begin{aligned}
\beta_{jt} &\sim N\left((\psi_{jt}/\psi_{j,t-1})^{1/2}\varphi_j\beta_{j,t-1}, (1-\varphi_j^2)\psi_{jt})\right) \\
\psi_{jt}|\kappa_{j,t-1} &\sim G\left(\lambda_j + \kappa_{j,t-1}, \frac{\lambda_j}{\gamma_j(1-\rho_j)}\right) \\
\kappa_{j,t-1}|\psi_{j,t-1} &\sim P\left(\frac{\rho_j\lambda_j\psi_{j,t-1}}{\gamma_j(1-\rho_j)}\right),
\end{aligned}$$

where $\beta_{j1} \sim N(0, \psi_{j1})$ and $\psi_{j,1} \sim G(\lambda_j, \lambda_j/\gamma_j)$.

ρ_j: autoregressive parameter for ψ_{jt}

γ_j: controls the overall relevance of β_{jt}

$$\begin{aligned}
E(\psi_{jt}|\psi_{j,t-1}) &= \gamma_j(1-\rho_j) + \rho_j\psi_{j,t-1} \\
V(\beta_{jt}) &= \gamma_j \quad \text{and} \quad \kappa(\beta_{jt}) = 3/\lambda_j
\end{aligned}$$

Vertical sparsity

Rocková and McAlinn [2018]
$\{\beta_1, \ldots, \beta_T\}$ follows a dynamic spike-and-slab (DSS) prior:

$$
\begin{aligned}
\beta_t | \beta_{t-1}, \theta_t \quad &\sim \quad (1-\theta_t)\left(\frac{0.9}{2}\exp\{-|\beta_t|0.9\}\right) \\
&+ \quad \theta_t \left(\frac{1}{\sqrt{2\pi(0.99)}}\exp\left\{-\frac{1}{2(0.99)}(\beta_t - 0.98\beta_{t-1})^2\right\}\right),
\end{aligned}
$$

where

$$
\theta_t = \frac{0.02\left(\frac{0.9}{2}\exp\{-|\beta_{t-1}|0.9\}\right)}{0.02\left(\frac{0.9}{2}\exp\{-|\beta_{t-1}|0.9\}\right) + 0.98\left(\frac{1}{\sqrt{2\pi(25)}}\exp\left\{-\frac{1}{2(25)}\beta_{t-1}^2\right\}\right)}
$$

They develop an optimization approach for dynamic variable selection.

Vertical sparsity

Kowal, Matteson, and Ruppert [2018]
As before, a given time-varying regression coefficient, β_t, is such that

$$\beta_t \sim N(0, \psi_t),$$

with dynamic shrinkage process

$$\log \psi_t = \mu + \phi(\log \psi_{t-1} - \mu) + \eta_t, \qquad \eta_t \sim Z(\alpha, \beta, 0, 1),$$

so

$$\begin{aligned} p(\eta_t) &\propto \{e^{\eta_t}\}^{\alpha}\{1 + e^{\eta_t}\}^{-(\alpha+\beta)} \\ e^{\eta_t} &\sim \text{inverted-Beta}. \end{aligned}$$

They name dynamic horseshoe process the above locally adaptive shrinkage.

Vertical sparsity: our proposal

Our contribution is defining as a spike-and-slab prior that not only shrinks time-varying coefficients in dynamic regression problems but allows for dynamic variable selection.

We use a non-centered parametrization:

$$\begin{aligned} y_t &= x_t' \tilde{\beta}_t + \nu_t, \qquad && \nu_t \sim N(0, \sigma_t^2) \\ \tilde{\beta}_t &= G_t \tilde{\beta}_{t-1} + \omega_t, \qquad && \omega_t \sim N(0, W_t), \end{aligned} \tag{6}$$

with $\beta_1 \sim N(0, \psi_1)$, where

$$\begin{aligned} \tilde{\beta}_t &= \left(\psi_{1t}^{-1/2} \beta_{1,t}, \ldots, \psi_{qt}^{-1/2} \beta_{qt}\right)' \\ G_t &= \text{diag}(\varphi_1, \ldots, \varphi_q) \\ W_t &= \text{diag}((1-\varphi_1^2), \ldots, (1-\varphi_q^2)), \\ x_t' &= (\psi_{1t}^{1/2} x_{1t}, \ldots, \psi_{qt}^{1/2} x_{qt}). \end{aligned}$$

Vertical sparsity: our proposal

We place independent priors for each $\psi_t = \tau^2 k_t$:

$$(k_t | k_{t-1} = \upsilon_i) \stackrel{\text{ind}}{\sim} (1 - \omega_{1i})\delta_r(\upsilon_i) + \omega_{1i}\delta_1(\upsilon_i),$$

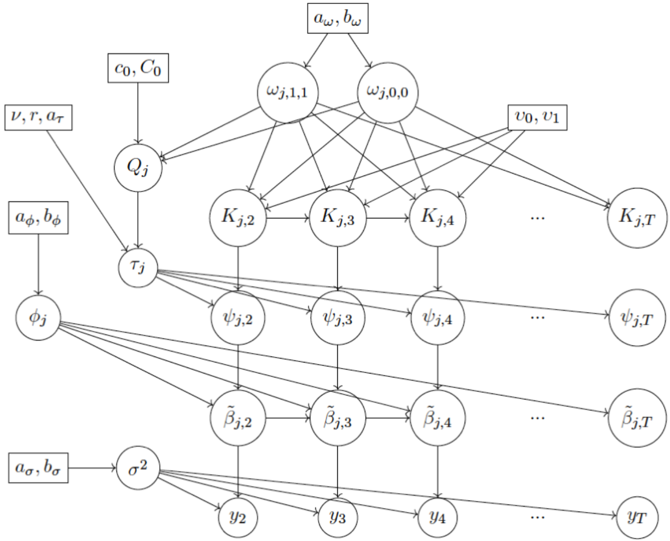
where $\upsilon_i \in \{r, 1\}$, $\omega_{1i} = p\left(k_t = 1 | k_{t-1} = \upsilon_i\right)$, and $p(k_1 = \upsilon_i) = 1/2$.

In addition,

$$r = \frac{\text{Var}_{spike}(\beta)}{\text{Var}_{slab}(\beta)} \ll 1,$$

and $p(\tau^2)$ is one of priors from the previous table.

Vertical sparsity: our proposal



MCMC sampling scheme I

Unsurprisingly, posterior inference is done via a customized MCMC scheme.

Full conditional distributions:

1. Draw k_ts using the algorithm of Gerlach, Carter, and Kohn [2000].
   - They proposed an efficient sampling algorithm, for dynamic mixture models, which samples k_t from $p(k_t|y_{1:T}, k_{s\neq t})$ without conditioning on $\tilde{\beta}_{1:T}$.
2. Draw $\tilde{\beta}_1, \ldots, \tilde{\beta}_T$, jointly and conditionally on $k_1, \ldots, k_T$, via forward filtering backward sampling (FFBS).
3. Draw σ^2:

$$(\sigma^2|\theta_{\setminus\sigma^2}, y) \sim IG\left(a_\sigma + \frac{T}{2}, b_\sigma + \frac{1}{2}\sum_{t=1}^{T}(y_t - x_t'\beta_t)^2\right).$$

MCMC sampling scheme II

4. Draw τ_j^2s:
   - Assuming the NMIG prior:

$$(\tau^2|\theta_{\setminus\tau^2}, y) \sim IG\left(\nu + \frac{T}{2}, Q + \frac{1}{2}\sum_{t=1}^{T}\frac{\left(\beta_t - k_t^{1/2}k_{t-1}^{-1/2}\varphi\beta_{t-1}\right)^2}{k_t(1-\varphi^2)}\right).$$

   - Assuming mixture of Laplaces or normal-gammas:

$$(\tau^2|\theta_{\setminus\tau^2}, y) \sim GIG(p, g, h),$$

   where

$$g = 1/Q, \quad h = \sum_{t=1}^{T}\frac{\left(\beta_t - k_t^{1/2}k_{t-1}^{-1/2}\varphi\beta_{t-1}\right)^2}{k_t(1-\varphi^2)}, \quad p = a_\tau - T/2.$$

MCMC sampling scheme III

5. Draw each φ_j using Metropolis Hastings step as its full conditional

$$p(\varphi_j|\theta_{\setminus\varphi_j}, y) \propto p(\varphi_j|a_\varphi, b_\varphi)p(\beta|k,\sigma^2,\tau^2,\varphi)$$
$$\propto \varphi_j^{(a_\varphi-1)}(1-\varphi_j)^{(b_\varphi-1)} \exp\left\{-\sum_{t=1}^{T}\frac{\left(\beta_t-\sqrt{\frac{\psi_t}{\psi_{t-1}}}\varphi_j\beta_{t-1}\right)^2}{2\psi_t(1-\varphi_j^2)}\right\},$$

has no close form. We use a Beta proposal density $q\left(\varphi_j^*|\varphi^{(m-1)}\right)$ as

$$\varphi_j^* \sim \mathcal{B}\left(\alpha, \xi\left(\varphi_j^{(m-1)}\right)\right), \quad \xi\left(\varphi_j^{(m-1)}\right) = \alpha\left(\frac{1-\varphi_j^{(m-1)}}{\varphi_j^{(m-1)}}\right),$$

where α is a tuning parameter and the acceptance distribution is

$$\mathcal{A}\left(\varphi_j^*|\varphi_j^{(m-1)}\right) = min\left\{1, \frac{p(\varphi_j^*|\theta_{\setminus\varphi_j}, y)q\left(\varphi_j^{(m-1)}|\varphi^*\right)}{p(\varphi_j^{(m-1)}|\theta_{\setminus\varphi_j}, y)q\left(\varphi_j^*|\varphi^{(m-1)}\right)}\right\}.$$

MCMC sampling scheme IV

6. Update the transition probabilities from the latent Markov process:

$$(\omega_{11}|\theta_{\backslash\omega_{11}}, y) \sim \mathcal{B}(a_\omega + \#\{t : \upsilon_1 \to \upsilon_1\}, b_\omega + \#\{t : \upsilon_1 \to \upsilon_0\}),$$
$$(\omega_{00}|\theta_{\backslash\omega_{00}}, y) \sim \mathcal{B}(a_\omega + \#\{t : \upsilon_0 \to \upsilon_0\}, b_\omega + \#\{t : \upsilon_0 \to \upsilon_1\}),$$

with $\upsilon_0 = r$ and $\upsilon_1 = 1$.

7. Drawing Q:

- NMIG prior: $GIG(p, g, h)$, with $p = \nu - c_0$, $g = 2\tau^{-2}$ and $h = 2[C_0/s^*]$,
- Mixture of Normal-Gammas: $IG(c_0 + a_\tau, \tau^2/2 + [C_0/s^*])$,
- Mixture of Laplaces: $IG(c_0 + a_\tau, \tau^2/2 + [C_0/s^*])$, with $a_\tau = 1$.

Outline

Example 1: Simulated dynamic regression

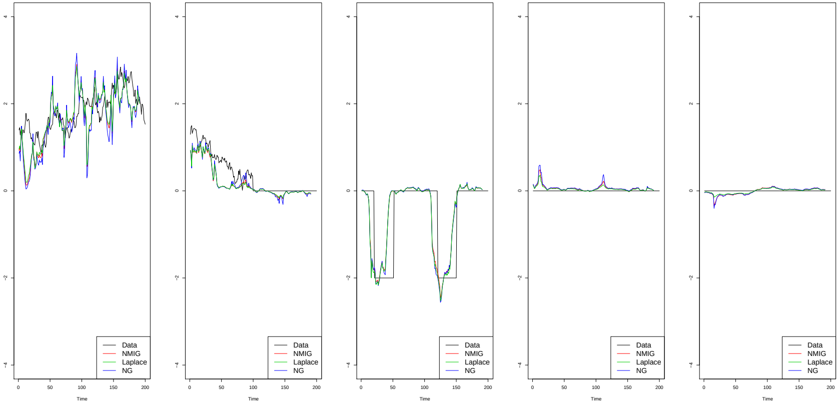


Figure 1: True β_t (black); posteriors means of β_t using dynamic NMIG (red), NG (blue) and Laplace (green) mixtures.

Example 2: Simulated Cholesky SV

We consider $T = 240$ and $q = 10$.

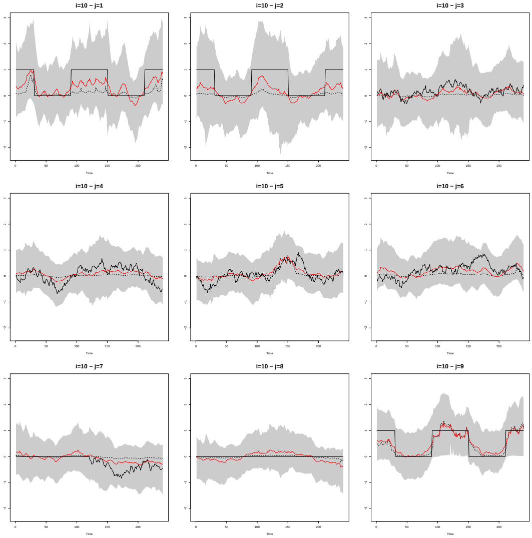


Figure 2: Coefficients for the last equation of the Cholesky recursive equations using dynamic Laplace prior.

Example 3: Inflation data

- Data: inflation data obtained from FRED database, Federal Reserve Bank of St.Louis, University of Michigan Consumer Survey database, Federal Reserve Bank of Philadelphia, and Institute of Supply Management with independent variable as the US quarterly inflation measure based on the Gross Domestic Product (GDP).
- The data includes 31 predictors, from activity and term structure variables to survey forecasts and previous lags. The sample period is from the second quarter of 1965 to first quarter of 2011 with $T = 182$ observations.
- The results (the mean of the coefficients $\beta_{j,t}$) were compared to results from the NGAR process defined from Kalli and Griffin [2014] (MATLAB code).

Inflation Data

Name	Description
GDP	Difference in logs of real gross domestic product
PCE	Difference in logs of real personal consumption expenditure
GPI	Difference in logs of real gross private investment
RGEGI	Difference in logs of real government consumption expenditure and gross investment
IMGS	Difference in logs of imports of goods and services
NFP	Difference in logs non-farm payroll
M2	Difference in logs M2 (commercial bank money)
ENERGY	Difference in logs of oil price index
FOOD	Difference in logs of food price index
MATERIALS	Difference in logs of producer price index (PPI) industrial commodities
OUTPUT GAP	Difference in logs of potential GDP level
GS10	Difference in logs of 10yr Treasury constant maturity rate
GS5	Difference in logs of 5yr Treasury constant maturity rate
GS3	Difference in logs 3yr Treasury constant maturity rate
GS1	Difference in logs 1yr Treasury constant maturity rate
PRIVATE EMPLOYMENT	Log difference in total private employment
PMI MANU	Log difference in PMI-manufacturing index
AHEPNSE	Log difference in average hourly earnings of private non management employees
DJIA	Log difference in Dow Jones Industrial Average Returns
M1	Log difference in M1 (narrow-commercial bank money)
ISM SDI	Institute for Supply Management (ISM) Supplier Deliveries Inventory
CONSUMER	University of Michigan consumer sentiment (level)
UNRATE	Log of the unemployment rate
TBILL3 3m	Treasury bill rate
TBILL SPREAD	Difference between GS10 and TBILL3
HOUSING STARTS	Private housing (in thousands of units)
INF EXP	University of Michigan inflation expectations (level)
LAG1, LAG2, LAG3, LAG4	The first, second, third and fourth lag

NGAR vs Dynamic SSL

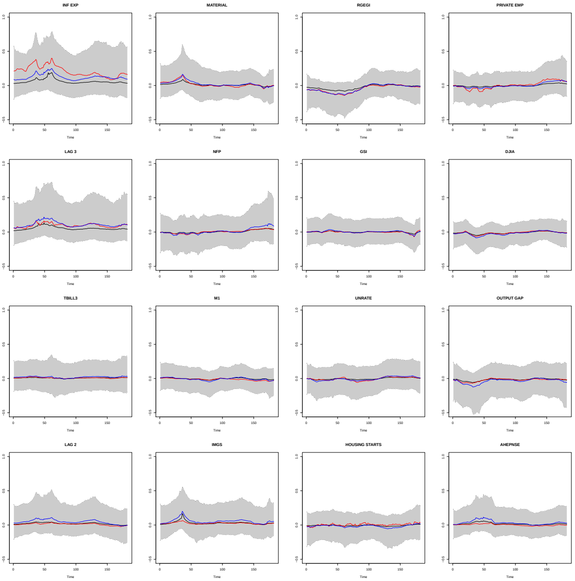


Figure 3: 16 most relevant predictors using dynamic NMIG prior.

Outline

Final remarks

- Test a Gamma prior for the parameter Q instead of the Inverted-Gamma or a Normal prior for $\sqrt{Q}$, following Frühwirth-Schnatter and Wagner [2010] which criticizes the use of the Inverse-Gamma because the posterior values are strongly influenced by the hyperparameters.
- Construct other mixture priors such as a dynamic mixture of Student's-t and Laplace densities for the coefficients.
- Allow for time-varying observational variances using stochastic volatility models.
- Allow for static coefficients besides time-varying coefficients as in Belmonte, Koop, and Korobilis [2014] and Bitto and Frühwirth-Schnatter [2016].
- Compare predictive performance with other existing methods.

References

Miguel AG Belmonte, Gary Koop, and Dimitris Korobilis. Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, 33(1):80–94, 2014.

Angela Bitto and Sylvia Frühwirth-Schnatter. Achieving shrinkage in a time-varying parameter model framework. *arXiv preprint arXiv:1611.01310*, 2016.

Carlos Carvalho, Hedibert Lopes, and Robert McCulloch. On the long run volatility of stocks. *Journal of the American Statistical Association (to appear)*, 2018.

Joshua CC Chan, Gary Koop, Roberto Leon-Gonzalez, and Rodney W Strachan. Time varying dimension models. *Journal of Business & Economic Statistics*, 30(3):358–367, 2012.

Sylvia Frühwirth-Schnatter and Hedibert F. Lopes. Sparse Bayesian factor analysis when the number of factors is unknown. Technical report, 2018.

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

Sylvia Frühwirth-Schnatter and Helga Wagner. Bayesian variable selection for random intercept modeling of gaussian and non-gaussian data. *Bayesian Statistics 9*, 9:165, 2011.

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

Richard Gerlach, Chris Carter, and Robert Kohn. Efficient bayesian inference for dynamic mixture models. *Journal of the American Statistical Association*, 95(451):819–828, 2000.

Jim Griffin and Philip Brown. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188, 2010.

Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

Hemant Ishwaran and J Sunil Rao. Spike and slab variable selection: frequentist and bayesian strategies. *Annals of Statistics*, pages 730–773, 2005.

Maria Kalli and Jim E Griffin. Time-varying sparsity in dynamic regression models. *Journal of Econometrics*, 178(2):779–793, 2014.

Gregor Kastner, Sylvia Frühwirth-Schnatter, and Hedibert F. Lopes. Efficient Bayesian inference for multivariate factor stochastic volatility models. *Journal of Computational and Graphical Statistics*, 26:905–917, 2017.

Gary Koop and Dimitris Korobilis. Large time-varying parameter VARs. *Journal of Econometrics*, 177:185–198, 2013.

Daniel R. Kowal, David S. Matteson, and David Ruppert. Dynamic shrinkage processes. Technical report, 2018.

F. B. Lempers. *Posterior Probabilities of Alternative Linear Models*. Rotterdam University Press, 1988.

Hedibert F. Lopes and Carlos M. Carvalho. Factor stochastic volatility with time varying loadings and markov switching regimes. *Journal of Statistical Planning and Inference*, 137:3082–3091, 2007.

Hedibert F. Lopes, Robert E. McCulloch, and Ruey S. Tsay. Parsimony inducing priors for large scale state-space models. Technical report, Booth School of Business, University of Chicago, 2008a.

Hedibert F. Lopes, Esther Salazar, and Dani Gamerman. Spatial dynamic factor models. *Bayesian Analysis*, 3:759–792, 2008b.

T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression (with discussion). *Journal of the American Statistical Association*, 83:1023–1036, 1988.

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

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

Veronika Rocková and Kenichiro McAlinn. Dynamic variable selection with spike-and-slab process priors. Technical report, Booth School of Business, University of Chicago, 2018.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.