Mixture Time Varying Parameter Models (=State Space Model plus Spike-and-slab)

Angela Bitto-Nemling
angela.bitto@wu.ac.at

Computational Statistics and Molecular Simulation:
A pratical cross-fertilization Oaxaca
12 November 2018

Mixture TVP Model

- BITTO (2017) Regularisation and Shrinkage Estimation for Dynamic Linear Models *Dissertation*.
- BITTO & FRÜHWIRTH-SCHNATTER (2018). Achieving Shrinkage in a TVP framework. *Journal of Econometrics*.
- BITTO & FRÜHWIRTH-SCHNATTER. Mixture TVP *Working paper*.

Cross-fertilization - please take this home from Oaxaca!

- Usefulness of switching parameterizations
- Shrinkage in state space models for the process variance
- Spike-and-slab priors - posterior inclusion probabilities
- Boosting by interweaving
- Algorithm available - R package out soon

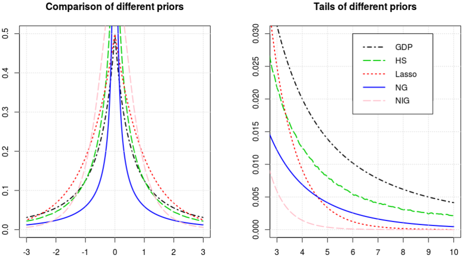


Figure 1 : Comparison of different shrinkage priors: generalized double Pareto prior, horseshoe prior, Bayesian Lasso, normal-gamma prior, normal-inverse Gaussian prior.

Centered State Space Model (TVP)

Observation equation:

$$y_t \quad = \quad \mathbf{x}_t \boldsymbol{\beta}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right),$$

State equation:

$$\boldsymbol{\beta}_t \quad = \quad \boldsymbol{\beta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}_d\left(\mathbf{0}, \mathbf{Q}\right),$$

with $\boldsymbol{Q} = \text{Diag}(\theta_1, \ldots, \theta_d)$.

Non-centered** state space model

Observation equation:

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + \mathbf{x}_t Diag\left(\sqrt{\theta_1}, \ldots, \sqrt{\theta_d}\right)\tilde{\boldsymbol{\beta}}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right),$$

State equations (independent), $j = 1, \ldots d$:

$$\tilde{\beta}_{jt} = \tilde{\beta}_{j,t-1,} + \tilde{\omega}_{jt}, \quad \tilde{\omega}_{jt} \sim \mathcal{N}\left(0, 1\right),$$

- $\boldsymbol{\beta}$ – initial value of the regression coefficients,
- $\theta_1, \ldots, \theta_d$ – process variances.

(Frühwirth-Schnatter and Wagner, 2010) (**Papaspiliopoulos, Gareth O. Roberts and Sköld (2007)).

Classification of parameters

$$\beta_{jt} = \beta_j + \sqrt{\theta_j}\tilde{\beta}_{jt}$$

β_j – initial value of the regression coefficients in the TVP model;
$\sqrt{\theta_1}, \ldots, \sqrt{\theta_d}$ – square root of the process variance.

- $\beta_j = 0$ & $\sqrt{\theta}_j = 0$ insignificant parameter
- $\beta_j \neq 0$ & $\sqrt{\theta}_j = 0$ constant parameter
- $\beta_j \neq 0$ & $\sqrt{\theta}_j \neq 0$ time varying parameter

Where to induce shrinkage?

$$y_t = \mathbf{x}_t \boldsymbol{\beta} + \mathbf{x}_t Diag\left(\sqrt{\theta_1}, \ldots, \sqrt{\theta_d}\right) \tilde{\boldsymbol{\beta}}_t + \varepsilon_t, \varepsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right),$$

$$\sqrt{\theta_j}|z_j \sim (1 - z_j) p_{spike}(\sqrt{\theta_j}|\eta) + z_j p_{slab}(\sqrt{\theta_j}|\eta),$$

$$\beta_j|z_j^* \sim (1 - z_j^*) p_{spike}(\beta_j|\eta^*) + z_j^* p_{slab}(\beta_j|\eta^*).$$

Spike-and-slab priors

- Useful prior in context of variable selection (Mitchell and Beauchamp, 1988, George and McCulloch 1993,1997)
- Finite mixture distribution with two components
- Use spike-and-slab smoothing priors for the process variances
- Posterior inclusion probabilities can be used for classification

$$r = \frac{\mathrm{Var}(\mathsf{spike}(\sqrt{\theta_j}))}{\mathrm{Var}(\mathsf{slab}(\sqrt{\theta_j}))} << 1.$$

$$\begin{aligned} p(z_j = 1|\omega) &= \omega, \\ p(\sqrt{\theta_j}|z_j) &= (1 - z_j)p_{spike}(\sqrt{\theta_j}) + z_j p_{slab}(\sqrt{\theta_j}). \end{aligned}$$

Normal-gamma prior in detail

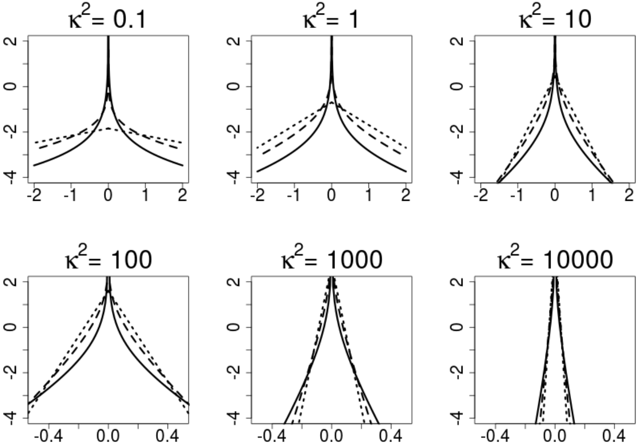


Figure 2 : Log of the prior density $\log p(\sqrt{\theta}_j|\kappa^2)$ for different values of κ^2 and $a^\xi = 0.1$ (solid line), $a^\xi = 1/3$ (dashed line) and $a^\xi = 1$ (dotted line).

Mixture TVP models - the essential slide

$$\pm\sqrt{\theta}_j|\xi_j^2 \quad \sim \quad \mathcal{N}\left(0, \xi_j^2\right), \qquad \xi_j^2|a^\xi, \kappa^2, z_j \sim \mathcal{G}\left(a^\xi, \frac{a^\xi \kappa^2}{2r(z_j)}\right),$$

with

$$r(z_j) = \begin{cases} 1 & \text{if } \; z_j = 1, \\ r << 1 & \text{if } \; z_j = 0, \end{cases}$$

Thus if $z_j = 1$ the prior variance is $\mathrm{Var}(\xi_j^2) = \frac{4}{a^\xi \kappa^4}$, while the prior variance $\mathrm{Var}(\xi_j^2) = \frac{4r^2}{a^\xi \kappa^4}$ decrease if $z_j = 0$.

Posterior inclusion probabilities

For classification we can thus work with the marginal densities

$$
\begin{aligned}
Pr(r(z_j) = 1 | \sqrt{\theta_j}, z_j) &\propto p(\sqrt{\theta_j} | \kappa^2, a^{\xi}, r(z_j) = 1), \\
Pr(r(z_j) = r | \sqrt{\theta_j}, z_j) &\propto p(\sqrt{\theta_j} | \kappa^2, a^{\xi}, r(z_j) = r),
\end{aligned}
$$

The marginal density $p(\sqrt{\theta}_j|\kappa^2)$, where ξ_j^2 is integrated out can be written as

$$
p(\sqrt{\theta}_j|\kappa^2) = \frac{\sqrt{a^{\xi}\kappa^2}^{a^{\xi}+1/2}}{\sqrt{\pi}2^{a^{\xi}-1/2}\Gamma(a^{\xi})} \frac{|\sqrt{\theta}_j|^{a^{\xi}-1/2}}{r(z_j)^{1/2(a^{\xi}+1/2)}} \mathcal{K}_{(a^{\xi}-1/2)} \left(\frac{\sqrt{a^{\xi}\kappa^2}|\sqrt{\theta_j}|}{\sqrt{r(z_j)}} \right)
$$

where $\mathcal{K}_p()$ is the modified Bessel function of the second kind with index p.

Algorithm (**TVP plus shrinkage**)

(a) *Sample all latent states* $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_0, \ldots, \tilde{\boldsymbol{\beta}}_T)$ $\tilde{\boldsymbol{\beta}}|\boldsymbol{\beta}, \mathbf{Q}, \sigma^2, \mathbf{P}_0 \sim \mathcal{N}_{(T+1)\cdot d}\left(\boldsymbol{\Omega}^{-1}\mathbf{c}, \boldsymbol{\Omega}^{-1}\right)$.

(b) *Joint sampling of* $\boldsymbol{\alpha} = (\beta_1, \ldots, \beta_d, \sqrt{\theta_1}, \ldots, \sqrt{\theta_d})'$ $\boldsymbol{\alpha}|\tilde{\boldsymbol{\beta}}, \boldsymbol{\tau}^2, \boldsymbol{\xi}, \sigma^2, \mathbf{y} \sim \mathcal{N}_{2d}\left(\mathbf{a}_T, \mathbf{A}_T\right)$.

(b*) *Boosting through ASIS. (Yu and Meng, (2011))*

(new-1) *Sample* z_j *from*

$$Pr(z_j = 1|\sqrt{\theta_j}, \omega) = \frac{1}{1 + \frac{1-\omega}{\omega} L_j}, \quad L_j = \frac{p_{spike}(\sqrt{\theta_j}|\kappa^2, a^{\xi}, z_j = 0)}{p_{slab}(\sqrt{\theta_j}|\kappa^2, a^{\xi}, z_j = 1)}.$$

(new-2) *Sample* ω *from* $\omega|z_1, \ldots, z_d \sim \mathcal{B}\left(a_\omega + n_1, b_\omega + d - n_1\right), \quad n_1 = \sum_{j=1}^{d} z_j$.

(c) *Sample the prior variances*

$$\begin{aligned} \xi_j^2|\theta_j, a^{\xi}, \kappa^2 &\sim \mathcal{GIG}\left(a^{\xi} - 1/2, a^{\xi}\kappa^2, \theta_j\right), \\ \tau_j^2|\beta_j, a^{\tau}, \lambda^2 &\sim \mathcal{GIG}\left(a^{\tau} - 1/2, a^{\tau}\lambda^2, \beta_j^2\right). \end{aligned}$$

(d) *Sample (for the homoscedastic model) the error variance* $\sigma^2|\mathbf{y}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}, C_0 \sim \mathcal{G}^{-1}\left(c_0 + \frac{T}{2}, C_0 + \frac{1}{2}\sum_{t=1}^{T}(y_t - \mathbf{z}_t\boldsymbol{\alpha})^2\right)$.

(e) *Sample the scale parameters of the initial distribution, from* $P_{0,jj}|\tilde{\beta}_{j0} \sim \mathcal{G}^{-1}\left(\nu_P + \frac{1}{2}, (\nu_P - 1)c_P + \frac{1}{2}\tilde{\beta}_{j0}^2\right)$.

Algorithm (**ASIS for TVP models**)

(b*-1) *Move to the centered parameterization; Store the sign of* $\sqrt{\theta_j}$;

$$\beta_{jt} = \beta_j + \sqrt{\theta_j}\tilde{\beta}_{jt}, \qquad t = 0, \ldots, T.$$

(b*-2) *Update* β_j *and* $\sqrt{\theta_j}$,

(b*-2a) *Draw* θ_j^{new} *from the conditional posterior* $\theta_j|\beta_j, \beta_{j0}, \ldots, \beta_{jT}, \xi_j^2, P_{0,jj} \sim \mathcal{GIG}\left(-\frac{T}{2}, \frac{1}{\xi_j^2}, \sum_{t=1}^{T}(\beta_{jt} - \beta_{j,t-1})^2 + \frac{(\beta_{j0}-\beta_j)^2}{P_{0,jj}}\right)$

(b*-2b) *Draw* β_j^{new} *from the conditional posterior* $\beta_j|\beta_{j0}, \theta_j^{\text{new}}, P_{0,jj}, \tau_j^2, \propto \mathcal{N}\left(\frac{\beta_{j0}\tau_j^2}{\tau_j^2+\theta_j P_{0,jj}}, \frac{\tau_j^2\theta_j P_{0,jj}}{\tau_j^2+\theta_j P_{0,jj}}\right)$.

(b*-2c) *Calculate* $\sqrt{\theta_j^{\text{new}}}$, *use the same (stored) sign as* $\sqrt{\theta_j}$.

(b*-3) *Move back to the non-centered:*

$$\tilde{\beta}_{jt}^{\text{new}} = (\beta_{jt} - \beta_j^{\text{new}})/\sqrt{\theta_j^{\text{new}}}, \qquad t = 0, \ldots, T.$$

ASIS - illustration

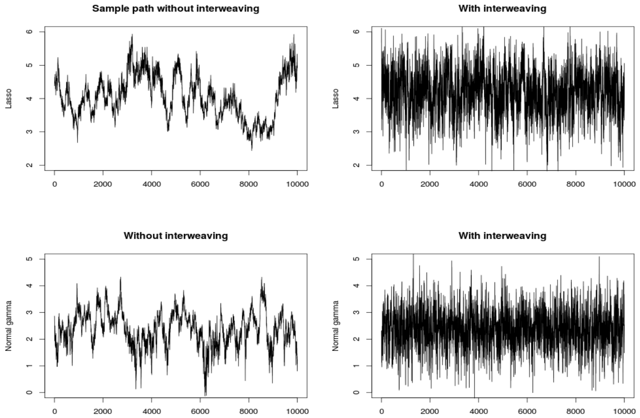


Figure 3 : ECB data. Sample paths comparing the MCMC schemes without interweaving $a^{\xi} = a^{\tau} = 1$ (top row) and $a^{\xi} = a^{\tau} = 0.2$ (bottom row). M =100,000 draws, only every tenth draw is shown.

Simulation study- visual inspection

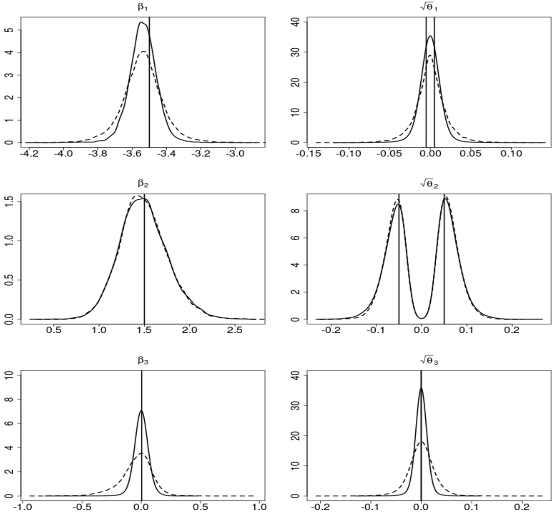


Figure 4 : Posterior densities of β_j and $\sqrt{\theta_j}$ together with the true values, $a^\tau = a^\xi = 0.1$ (solid line) and $a^\tau = a^\xi = 1$ (dashed line).

Simulation study- visual inspection

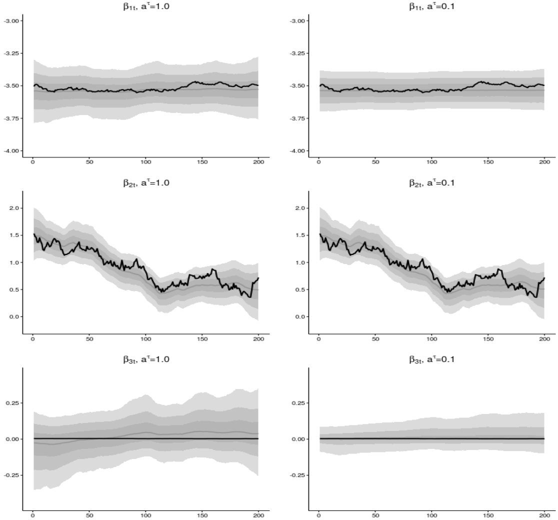


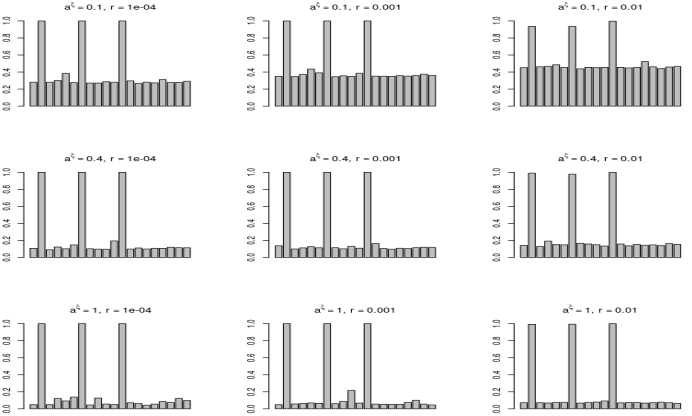
Figure 5 : Pointwise quantiles of the posterior paths $\beta_{jt} = \beta_j + \sqrt{\theta_j}\tilde{\beta}_{jt}$; left hand side: $a^\tau = a^\xi = 1$, right hand side: $a^\tau = a^\xi = 0.1$.

Simulation study

	$a^\tau = a^\xi = 0.1$			$a^\tau = a^\xi = 1$		
Name	$avMSE$	$avVAR$	$avBIAS^2$	$avMSE$	$avVAR$	$avBIAS^2$
β_1	7.7E-03	6.9E-03	7.7E-04	1.6E-02	1.5E-02	9.4E-04
β_2	1.8E-01	6.0E-02	1.2E-01	1.5E-01	5.5E-02	9.3E-02
β_3	2.9E-03	2.5E-03	4.0 E-04	2.7E-02	2.0E-02	6.5E-03
$\lvert\sqrt{\theta_1}\rvert$	6.9E-05	6.8E-05	1.7E-06	2.7E-04	1.9E-04	8.3E-05
$\lvert\sqrt{\theta_2}\rvert$	9.0E-04	6.2E-04	2.8E-04	6.4E-04	4.9E-04	1.5E-04
$\lvert\sqrt{\theta_3}\rvert$	1.5 E-04	1.1 E-04	3.8E-05	7.0E-04	3.0E-04	4.0E-04

Table 1 : Simulated data. Average mean squared error ($avMSE$), average variance ($avVAR$), and average squared bias ($avBIAS^2$) over 100 simulated time series.

Simulation study - posterior inclusion probabilities



Application - EU-area inflation modeling

- EU area inflation modeling based on generalized Phillips curve Belmonte et al. (2014))

$$y_{t+h} = c_t + \sum_{j=0}^{p-1} \phi_{j,t} \cdot y_{t-j} + \boldsymbol{\alpha}_t \mathbf{x}_t + s_t + \varepsilon_{t+h}$$

- Inflation depends on 37 possibly time-varying coefficients
  - lags of inflation (p=12)
  - 11 monthly seasonal dummies
  - 13 other predictors $\mathbf{x}_t$ (1-month Euribor, 1-year Euribor, Δ in industrial production index, Δ in monetary aggregate M3, Unemployment rate, Δ oil price)
- Monthly data from February 1994 until November 2010 $(t = 190)$

Classification based on visual inspection

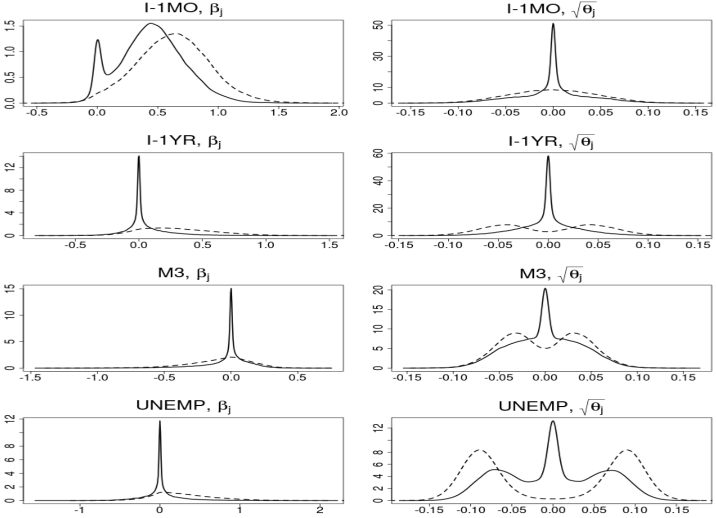


Figure 6 : ECB data. Posterior densities of β_j (left hand side) and $\sqrt{\theta_j}$ (right hand side)

Classification based on visual inspection

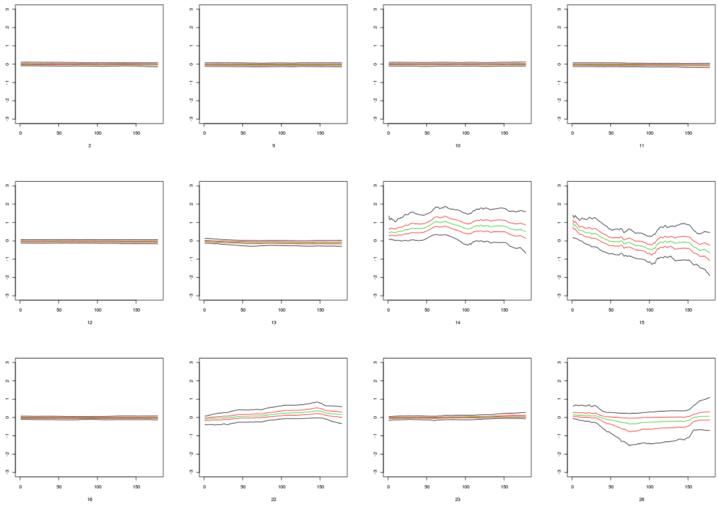


Figure 7 : ECB data. Quantiles of the posterior paths of $\beta_{jt} = \beta_j + \sqrt{\theta_j}\tilde{\beta}_{jt}$

Classification based on posterior inclusion probabilities

Figure 8 : ECB data. Posterior inclusion probabilities of $\sqrt{\theta_j}$

Interweaving matters

	$a^\tau = a^\xi = 0.2$				$a^\tau = a^\xi = 1$			
	no ASIS		ASIS		no ASIS		ASIS	
j	β_j	$\lvert\sqrt{\theta_j}\rvert$	β_j	$\lvert\sqrt{\theta_j}\rvert$	β_j	$\lvert\sqrt{\theta_j}\rvert$	β_j	$\lvert\sqrt{\theta_j}\rvert$
1	3641	211	133	94	1749	158	97	67
14	248	271	82	149	84	110	38	83
15	101	240	70	134	110	135	55	76
22	172	220	64	140	119	96	45	63
26	630	412	89	238	485	158	63	55

Table 2 : ECB data. Inefficiency factors of MCMC posterior draws of selected parameters with and without interweaving under shrinkage priors with $d_1 = d_2 = e_1 = e_2 = 1$ and, respectively, $a^\tau = a^\xi = 0.2$ and $a^\tau = a^\xi = 1$.

Predictive Evaluation for EU-area inflation modelling

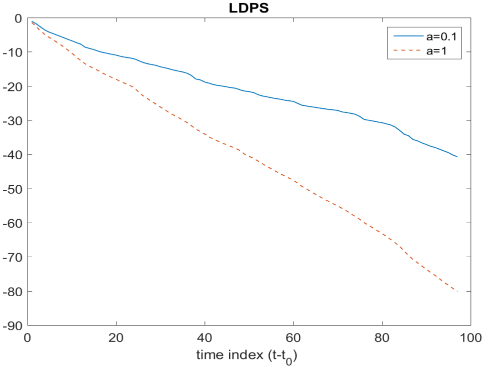


Figure 9 : ECB data. Cumulative log predictive scores for the last 100 points in time obtained through the conditionally optimal Kalman mixture approximation under $a^\tau = a^\xi = 0.1$ (solid line) and $a^\tau = a^\xi = 1$ (dashed line) with $d_1 = d_2 = e_1 = e_2 = 0.001$.

Predictive Evaluation for EU-area inflation modelling

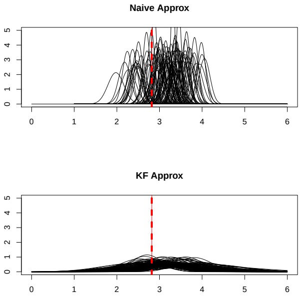


Figure 10 : ECB data. Illustration of the naive mixture approximation and the conditionally optimal Kalman mixture approximation for the predictive density at $t = t_0 + 40$ for $a^\tau = a^\xi = 0.1$ (left hand side) and $a^\tau = a^\xi = 1$ (right hand side) for $d_1 = d_2 = e_1 = e_2 = 0.001$.

Conclusion - what to take away?

- Introduction
  - Discuss different shrinkage priors
- Introduce shrinkage for the process variances θ_j through normal-gamma prior.
- Propose an efficient MCMC scheme for TVP models with shrinkage
  - Boost this algorithm with an interweaving strategy
- Discuss the approximation of the one-step ahead predictive densities
  - *Conditionally optimal Kalman filter approximation*
  - *Naive Gaussian mixture approximation*.
- Applications - simulations study, univariate data
- Mixture TVP model

Generalized inverse Gaussian distribution

The conditionally normal prior $\sqrt{\theta_j} \sim \mathcal{N}\left(0, \xi_j^2\right)$ leads to a posterior of $\xi_j^2|\sqrt{\theta_j}$ where the likelihood is the kernel of an inverted Gamma density in ξ_j^2 which is combined with the Gamma prior $\xi_j^2 \sim \mathcal{G}\left(a^\xi, a^\xi\kappa^2/2\right)$ - this leads to the GIG.

Generalized inverse Gaussian distribution

$Y \sim \mathcal{GIG}(p, a, b), a > 0, b > 0$ is a three parameter family with support on $y \in \mathbb{R}^+$.

$$p(y) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})} y^{p-1} e^{-(a/2)y} e^{-b/(2y)},$$

and $K_p(z)$ is the modified Bessel function of the second kind.
Very useful implementation: [Hörmann & Leydold (2015)].

Forecasting details

$$\begin{aligned} \mathrm{LPDS} = \log p(y_{t_0+1}, \ldots, y_T | \mathbf{y}^{\mathrm{tr}}) &= \\ \sum_{t=t_0+1}^{T} \log p(y_t | y^{t-1}) &= \sum_{t=t_0+1}^{T} \mathrm{LPDS}_t^{\star}, \end{aligned}$$

[Good (1952)], [Diebold et al. (1998)],[Diks et al. (2011)][Gneiting (2011)],[Gneiting and Ranjan (2011)],[Geweke and Keane (2007)],[Villani et al. (2009)]

Bibliography I

Armagan, A., Dunson, D., & Clyde, M., (2011). Generalized beta mixtures of Gaussians. *Advances in Neural Information Processing Systems 24*, 523-531.

Armagan, A., Dunson, D., & Lee, J., (2013). Generalized double Pareto shrinkage. *Statistica Sinica 23*, 119-143.

Amisano, G., & Geweke, J., (2013). Prediction using several macroeconomic models. *Working Paper Series 1537, European Central Bank*.

Andrews, D. F., & Mallows, C. L. (1974). Scale Mixture of Normal Distributions. *Journal of the Royal Statistical Society. Series B (Methodological) Vol. 36, No. 1* , pp. 99-102.

Bibliography II

- Belmonte, M. A. G, Koop, G. & Korobilis, D. (2013). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting, Vol. 33, Issue 1*, 80-94.
- Carvalho, C. M., & Polson, N. G., (2010). The horseshoe estimator for sparse signals. *Biometrika 97 (2)*, 465-480.
- Dangl, T. & Halling, M. (2009). Predictive Regressions with Time-Varying Coefficients. *Journal of Financial Economics, Vol. 106, Issue 1*, 157-181.
- Durbin, J., & Koopmann, S. J., (2001). Time Series Analysis by State Space Methods. Oxford: Oxford University Press.
- Geweke

Bibliography III

- FAHRMEIER, L., KNEIB, T. & KONRATH S. (2010). Bayesian regularisation and in structured additive regressions: A unifying perspective on shrinkage, smoothing and predictor selection. *Statist. Computing 20*, 203-219.
- FRANK, I. E., & FRIEDMAN, J. H., (1993) A Statistical View of Some Chemometrics Regression Tools. *Technometrics, Vol. 35, No. 2*, 109-148.
- GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis 6*, 515-533.
- GRIFFIN, J.E. & BROWN P. J., (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis 5*, 171-188.

Bibliography IV

- Griffin, J.E. & Brown P. J., (2011). Bayesian hyper lassos with non-convex penalisation. *Australian and New Zealand Journal of Statistics 53*, 423-442.
- Kalli, M., & Griffin, J. E., (2014) Time-Varying Sparsity in Dynamic Regression Models. *Journal of Econometrics, Vol. 178, Issue 2*, 779-793.
- Good, I. J., (1952), Rational Decisions. *Journal of the Royal Statistical Society, Series B 14*, 107-114.
- Hörmann W. & Leydold J., (2013). Generating generalized inverse Gaussian random variates. *Computational Statistics & Data Analysis 55 (1)*, 213-217.
- Hörmann W. & Leydold J., (2015). Package âĂŸGIGrvgâĂŹ. *XX*.

Bibliography V

- Hoerl, A. E. & Kennard, R. W., (1970) Ridge Regression: Biased estimation for Nonorthogonal Problems. *Technometrics, Vol. 12, No. 1*, 55-67.
- Kastner G. & Frühwirth-Schnatter, S., (2014). Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Estimation of Stochastic Volatility Models. *Computational Statistics and Data Analysis 76*, 408-423.
- Kastner G. (20XX). R-package *Text*
- Carter, C. K., & Kohn, R., (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika 83*, 589-601.
- Kim, S., Shephard, N. & Chip, S., (1998). Stochastic volatility: likelihood inference and comparison with ARCH models *Review of Economic Studies 65*, 361-393.

Bibliography VI

- Li, Q., & Lin, N. (2010). The Bayesian Elastic Net. *Bayesian Analysis 5 (1)*, 151-170.
- Lopes, H., F., & McCulloch, R., E., (2013). Generating generalized inverse Gaussian random variates. *Computational Statistics & Data Analysis 55 (1)*, 213-217.
- McCausland, W. J., Miller S. & Pelletier D., (2011). Simulation smoothing for state space models: A computational efficiency analysis. *Computational Statistics & Data Analysis 55*, 199-212.
- Nakajima, J., (2011). Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications. *Monetary and Economic Studies 29*, 107-142.

Bibliography VII

- Nakajima, J. & West, M., (2013) Bayesian analysis of latent threshold dynamic models. *Journal of Business and Economic Statistics 31*, 151-164.
- O'Hara, R.B. & Sillanpää, M. J., (2009). A review of Bayesian Variable Selection Methods: What, How and Which. *Bayesian Analysis 4*, 85-118.
- Polson, N. G., Scott J. G., & Windle J. (2013). The Bayesian bridge. *Journal of the Royal Statistical Society, Series B, 76*, 713-733.
- Park T., & Casella, G., (2008). The Bayesian Lasso. *Journal of the American Statistical Association 103*, 681-686.
- Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies 72* , 821âĂŞ852.

Bibliography VIII

- FRÜHWIRTH-SCHNATTER, S., (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis 15*, 183-202.
- FRÜHWIRTH-SCHNATTER, S., (2004). Computationally efficient Bayesian parameter estimation for State Space Models based on Reparametrizations *XX*, XX.
- STOCK J., & WATSON M. (2011). Generalized shrinkage methods for forecasting using many predictors. *XX*, XX.
- SIMS C. (2001). Comment on Cogley and Sargent (2001) *NBER Macroeconomics Annual*,Edited by B.S.Bernanke and K. Rogoff, Cambridge, MA:MIT Press, 373-379.

Bibliography IX

- Frühwirth-Schnatter, S. & Wagner H., (2010). Stochastic model specification search for Gaussian and partially non-Gaussian state space models. *Journal of Econometrics 154*, 85-100.
- Frühwirth-Schnatter, S. & Wagner H., (2011). Bayesian variable selection for random intercept modeling of Gaussian and non-Gaussian data. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckmann, A. F. M. Smith, & M. West (Eds.), *Bayesian Statistics 9*, 165-200. Oxford University Press.
- Tibshirani R., (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Ser. B 58*, 267-288.

Bibliography X

- Wagner H. & Duller C., (2011). Bayesian model selection for logistic regression models with random intercept. *Computational Statistics and Data Analysis*, published online.
- West M., & Harrison, J., (1999) Bayesian Forecasting and Dynamic Models.
- Armagan, A., Dunson, D., & Clyde, M., (2011). Generalized beta mixtures of Gaussians. *Advances in Neural Information Processing Systems 24*, 523-531.
- Armagan, A., Dunson, D., & Lee, J., (2013). Generalized double Pareto shrinkage. *Statistica Sinica 23*, 119-143.
- Amisano, G., & Geweke, J., (2013). Prediction using several macroeconomic models. *Working Paper Series 1537, European Central Bank*.

Bibliography XI

- ANDREWS, D. F., & MALLOWS, C. L. (1974). Scale Mixture of Normal Distributions. *Journal of the Royal Statistical Society. Series B (Methodological) Vol. 36, No. 1* , pp. 99-102.
- BELMONTE, M. A. G, KOOP, G. & KOROBILIS, D. (2013). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting, Vol. 33, Issue 1*, 80-94.
- CARVALHO, C. M., & POLSON, N. G., (2010). The horseshoe estimator for sparse signals. *Biometrika 97 (2)*, 465-480.
- DANGL, T. & HALLING, M. (2009). Predictive Regressions with Time-Varying Coefficients. *Journal of Financial Economics, Vol. 106, Issue 1*, 157-181.

Bibliography XII

- DURBIN, J., & KOOPMANN, S. J., (2001). Time Series Analysis by State Space Methods. Oxford: Oxford University Press.
- Geweke
- FAHRMEIER, L., KNEIB, T. & KONRATH S. (2010). Bayesian regularisation and in structured additive regressions: A unifying perspective on shrinkage, smoothing and predictor selection. *Statist. Computing 20*, 203-219.
- FRANK, I. E., & FRIEDMAN, J. H., (1993) A Statistical View of Some Chemometrics Regression Tools. *Technometrics, Vol. 35, No. 2*, 109-148.
- GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis 6*, 515-533.

Bibliography XIII

- Griffin, J.E. & Brown P. J., (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis 5*, 171-188.
- Griffin, J.E. & Brown P. J., (2011). Bayesian hyper lassos with non-convex penalisation. *Australian and New Zealand Journal of Statistics 53*, 423-442.
- Kalli, M., & Griffin, J. E., (2014) Time-Varying Sparsity in Dynamic Regression Models. *Journal of Econometrics, Vol. 178, Issue 2*, 779-793.
- Good, I. J., (1952), Rational Decisions. *Journal of the Royal Statistical Society, Series B 14*, 107-114.
- Hörmann W. & Leydold J., (2013). Generating generalized inverse Gaussian random variates. *Computational Statistics & Data Analysis 55 (1)*, 213-217.

Bibliography XIV

- Hörmann W. & Leydold J., (2015). Package âĂŸGIGrvgâĂŹ. *XX*.
- Hoerl, A. E. & Kennard, R. W., (1970) Ridge Regression: Biased estimation for Nonorthogonal Problems. *Technometrics, Vol. 12, No. 1*, 55-67.
- Kastner G. & Frühwirth-Schnatter, S., (2014). Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Estimation of Stochastic Volatility Models. *Computational Statistics and Data Analysis 76*, 408-423.
- Kastner G. (20XX). R-package *Text*
- Carter, C. K., & Kohn, R., (1996). On Gibbs sampling for state-space models. *Biometrika 81*, 541-553.

Bibliography XV

- CARTER, C. K., & KOHN, R., (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika 83*, 589-601.
- KIM, S., SHEPHARD, N. & CHIP, S., (1998). Stochastic volatility: likelihood inference and comparison with ARCH models *Review of Economic Studies 65*, 361-393.
- LI, Q., & LIN, N. (2010). The Bayesian Elastic Net. *Bayesian Analysis 5 (1)*, 151-170.
- LOPES, H., F., & MCCULLOCH, R., E., (2013). Generating generalized inverse Gaussian random variates. *Computational Statistics & Data Analysis 55 (1)*, 213-217.

Bibliography XVI

- McCausland, W. J., Miller S. & Pelletier D., (2011). Simulation smoothing for state space models: A computational efficiency analysis. *Computational Statistics & Data Analysis 55*, 199-212.
- Nakajima, J., (2011). Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications. *Monetary and Economic Studies 29*, 107-142.
- Nakajima, J. & West, M., (2013) Bayesian analysis of latent threshold dynamic models. *Journal of Business and Economic Statistics 31*, 151-164.
- O'Hara, R.B. & Sillanpää, M. J., (2009). A review of Bayesian Variable Selection Methods: What, How and Which. *Bayesian Analysis 4*, 85-118.

Bibliography XVII

- Polson, N. G., Scott J. G., & Windle J. (2013). The Bayesian bridge. *Journal of the Royal Statistical Society, Series B, 76*, 713-733.
- Park T., & Casella, G., (2008). The Bayesian Lasso. *Journal of the American Statistical Association 103*, 681-686.
- Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies 72* , 821âĂŞ852.
- Frühwirth-Schnatter, S., (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis 15*, 183-202.
- Frühwirth-Schnatter, S., (2004). Computationally efficient Bayesian parameter estimation for State Space Models based on Reparametrizations *XX*, XX.

Bibliography XVIII

- STOCK J., & WATSON M. (2011). Generalized shrinkage methods for forecasting using many predictors. *XX*, XX.
- SIMS C. (2001). Comment on Cogley and Sargent (2001) *NBER Macroeconomics Annual*,Edited by B.S.Bernanke and K. Rogoff, Cambridge, MA:MIT Press, 373-379.
- FRÜHWIRTH-SCHNATTER, S. & WAGNER H., (2010). Stochastic model specification search for Gaussian and partially non-Gaussian state space models. *Journal of Econometrics 154*, 85-100.
- FRÜHWIRTH-SCHNATTER, S. & WAGNER H., (2011). Bayesian variable selection for random intercept modeling of Gaussian and non-Gaussian data. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckmann, A. F. M. Smith, & M. West (Eds.), *Bayesian Statistics 9*, 165-200. Oxford University Press.

Bibliography XIX

- TIBSHIRANI R., (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Ser. B 58*, 267-288.
- WAGNER H. & DULLER C., (2011). Bayesian model selection for logistic regression models with random intercept. *Computational Statistics and Data Analysis*, published online.
- WEST M., & HARRISON, J., (1999) Bayesian Forecasting and Dynamic Models.