Time-Varying Parameter Models –
Efficient Bayesian Estimation using Shrinkage Priors

Sylvia Frühwirth-Schnatter and Angela Bitto
Department of Finance, Accounting and Statistics
Vienna University of Economics and Business, Austria

Time-varying parameter (TVP) models

TVP models very popular in recent work in applied econometrics:

- Primiceri (2005): time-varying structural VARs in a monetary policy application
- Dangl and Halling (2012): TVP model for equity return prediction
- Lopes et al. (2012): time-varying multivariate SV model for high-dimensional financial time series

Time-varying parameter (TVP) models

- Belmonte et al. (2014): TVP model with many predictors to forecast EU-area inflation

- Eisenstat et al. (2014): Time-Varying Parameter VARs with an application to analyze the response of macro variables to fiscal shocks

. . . and many others

Why should we care about shrinkage?

- Time-varying parameter models are very popular because of their flexibility, however ...

- Allowing flexibility where none is required may lead to
  - **inefficient** estimation for parameters of interest
  - **poor predictions**

- Shrinkage: allow flexibility where needed, but simplify wherever possible.

How to achieve shrinkage?

How shrinkage is introduced:

- Non-Bayesian: use regularization

- Bayesian: substitute "usual priors" by shrinkage priors that introduce sparsity.

There is a close connection between regularization and shrinkage priors.

Regularization and Log Posteriors in Regression Analysis

Regression model: $y_i = \mathbf{x}_i\boldsymbol{\alpha} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$.

LASSO (Tibshirani, 1996) - penalized sum of squared errors:

$$\sum_{i=1}^{N}(y_i - \mathbf{x}_i\boldsymbol{\alpha})^2 + \lambda \sum_{j=1}^{d} |\alpha_j|$$

Bayesian LASSO (Park and Casella, 2008) - log posterior $\log p(\boldsymbol{\alpha}|\mathbf{y}, \sigma_\varepsilon^2, \tau)$ under prior $\alpha_j \sim \mathrm{Lap}\left(0, \sigma_\varepsilon^2\tau\right)$:

$$-\frac{0.5}{\sigma_\varepsilon^2}\left(\sum_{i=1}^{N}(y_i - \mathbf{x}_i\boldsymbol{\alpha})^2 + \frac{1}{\tau}\sum_{j=1}^{d}|\alpha_j|\right) + \text{constant}$$

Regularization and Log Posteriors

- Any penalty $\psi(\boldsymbol{\alpha})$ in likelihood estimation translates into a prior $p(\boldsymbol{\alpha})$ and vice versa; see Fahrmeir et al. (2010).

- No messing around with the likelihood $p(\mathbf{y}|\boldsymbol{\alpha})$; sparsity introduced through the prior (penalty).

- **Principled Bayesian framework** to introduce sparsity into latent variable models: random coefficient models (Frühwirth-Schnatter and Tüchler, 2008), fixed-effects modeling of Gaussian and non-Gaussian panel data (Frühwirth-Schnatter and Wagner, 2011; Wagner and Duller, 2012), unobserved components models (Frühwirth-Schnatter and Wagner, 2010), . . .

Sparsity/shrinkage/variable selection for TVP models

Let y_t, $t = 1, \ldots, T$, be time series observations, which are supposed to be driven by latent variables, summarized in a state vector $\boldsymbol{\beta}_t$ we are unable to observe.

Observation equation:

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

State equation for the time-varying regression parameter:

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

Sparsity/shrinkage/variable selection for TVP models

Assume that $\mathbf{Q} = \mathrm{Diag}\,(\theta_1, \ldots, \theta_r)$, i.e.:

$$\beta_{tj} = \beta_{t-1,j} + w_{jt}, \qquad w_{jt} \sim \mathcal{N}\left(0, \theta_j\right), \quad j = 1, \ldots, r,$$
$$y_t = \mathbf{x}_t \boldsymbol{\beta}_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2_\varepsilon\right),$$

with initial values $\beta_{0,j} \sim \mathcal{N}\left(\beta_j, \theta_j P_{0j}\right)$.

- Why introduce shrinkage?
- How to introduce shrinkage?

Sparsity/shrinkage/variable selection for TVP models

Shrinkage priors should allow to discriminate between the following cases:

- $\beta_j \neq 0$ and $\theta_j \neq 0$, in which case $\beta_{tj} \neq \beta_{t-1,j}$ for all $t = 1, \ldots, T$
  $\Rightarrow$ **Coefficient dynamic (time-varying)**

- $\beta_j \neq 0$ and $\theta_j = 0$, in which case $\beta_{tj} \equiv \beta_j$ for all $t = 1, \ldots, T$
  $\Rightarrow$ **Coefficient significant, but static**

- $\beta_j = 0$ and $\theta_j = 0$, in which case $\beta_{tj} = 0$ for all $t = 1, \ldots, T$
  $\Rightarrow$ **Coefficient insignificant**

Shrinkage priors for variances in state space models

Shrinkage priors should allow the posterior to concentrate over the reduced model, if the model is overfitting:

- The standard choice for the prior of the variances is the conditional conjugate prior, i.e. $\theta_j \sim \mathcal{G}^{-1}(c_0, C_0)$. Bounds the posterior away from 0.

- Useful shrinkage priors: normal prior on the signed square root, $\pm\sqrt{\theta_j} \sim \mathcal{N}(0, B_0)$ (Frühwirth-Schnatter, 2004; Frühwirth-Schnatter and Wagner, 2010); equivalent to:

$$\theta_j \sim \mathcal{G}(1/2, 1/(2B_0)) = B_0 \cdot \chi^2_1.$$

Unobserved component models

UC model with a local level:

$$\mu_t = \mu_{t-1} + \omega_{1t}, \qquad \omega_{1t} \sim \mathcal{N}\left(0, \theta_1\right),$$
$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2_\varepsilon\right)$$

UC model with a local trend:

$$\mu_t = \mu_{t-1} + a_{t-1} + \omega_{1t}, \qquad \omega_{1t} \sim \mathcal{N}\left(0, \theta_1\right)$$
$$a_t = a_{t-1} + \omega_{2t}, \qquad \omega_{2t} \sim \mathcal{N}\left(0, \theta_2\right),$$
$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2_\varepsilon\right)$$

UC model with a local level

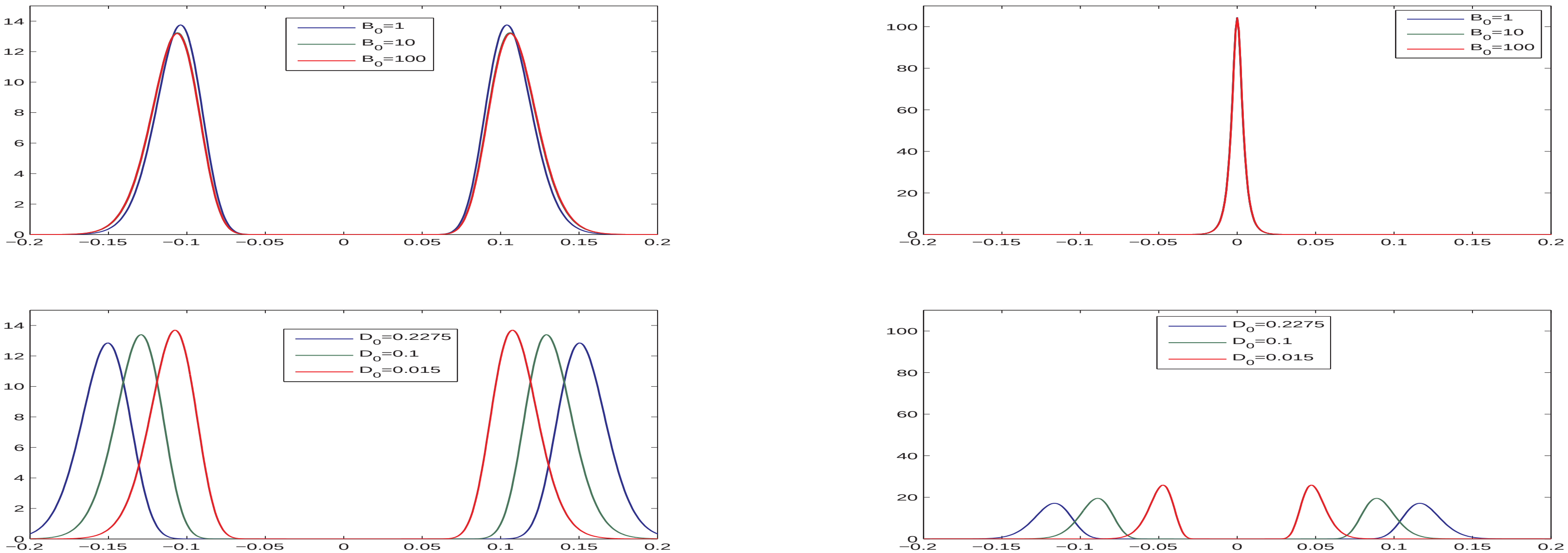


Figure 1: Posterior density for $\pm\sqrt{\theta}$ (data simulated from an UC model with local level, $\sigma_\varepsilon^2 = 1$, $T = 100$); left: $\theta = 0.01$; right: $\theta = 0$; top: prior $\theta \sim \mathcal{G}(0.5, 0.5/B_0)$, bottom: prior $\theta \sim \mathcal{G}^{-1}(0.5, C_0)$

UC model with a local trend

If the true variances θ_1^{true} and θ_2^{true} are positive, then the likelihood function concentrates around four modes.

Figure 2: Contour and surface plots of the (scaled) profile likelihood $l(\sqrt{\theta_1}, \sqrt{\theta_2}) = p(\mathbf{y}|\sqrt{\theta_1}, \sqrt{\theta_2}, \sigma_\varepsilon^{2,\text{true}}, \mu_0^{\text{true}}, a_0^{\text{true}})$ for simulated data $(T = 1000)$ with $(\theta_1^{\text{true}}, \theta_2^{\text{true}}) = (0.15, 0.02)$

UC model with a local trend

If one the true variances θ_1^{true} and θ_2^{true} is equal to 0 while the other is positive, two of those modes collapse and the likelihood is bimodal with an increasing number of observations T.

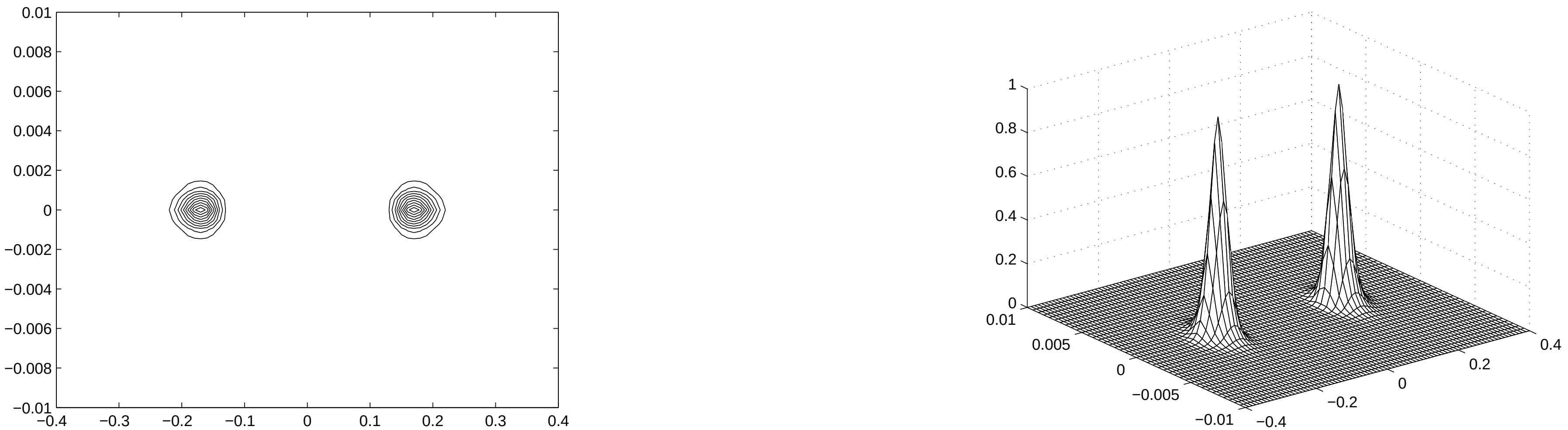
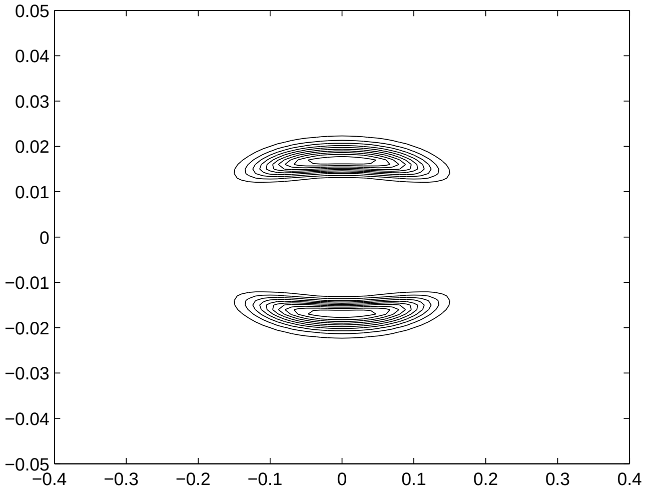


Figure 3: Contour and surface plots of the (scaled) profile likelihood for simulated data ($T = 1000$) with $(\theta_1^{\text{true}}, \theta_2^{\text{true}}) = (0.15, 0)$

UC model with a local trend



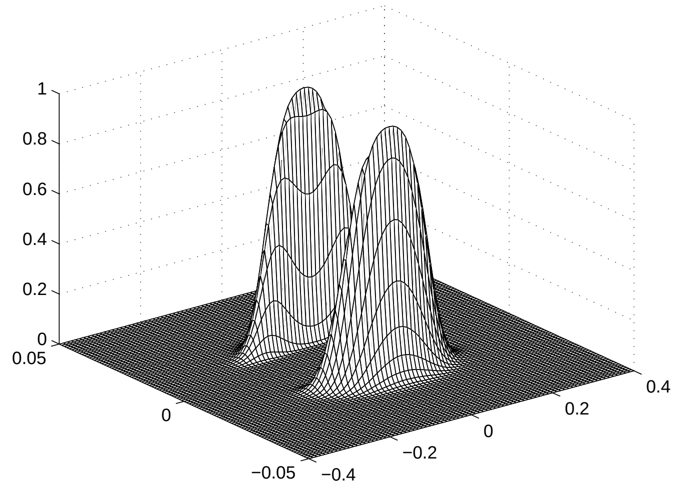


Figure 4: Contour and surface plots of the (scaled) profile likelihood for simulated data $(T = 1000)$ with $(\theta_1^{\mathrm{true}}, \theta_2^{\mathrm{true}}) = (0, 0.02)$

UC model with a local trend

If both variances θ_1^{true} and θ_2^{true} are equal to zero, then the likelihood function will be unimodal with an increasing number of observations T.

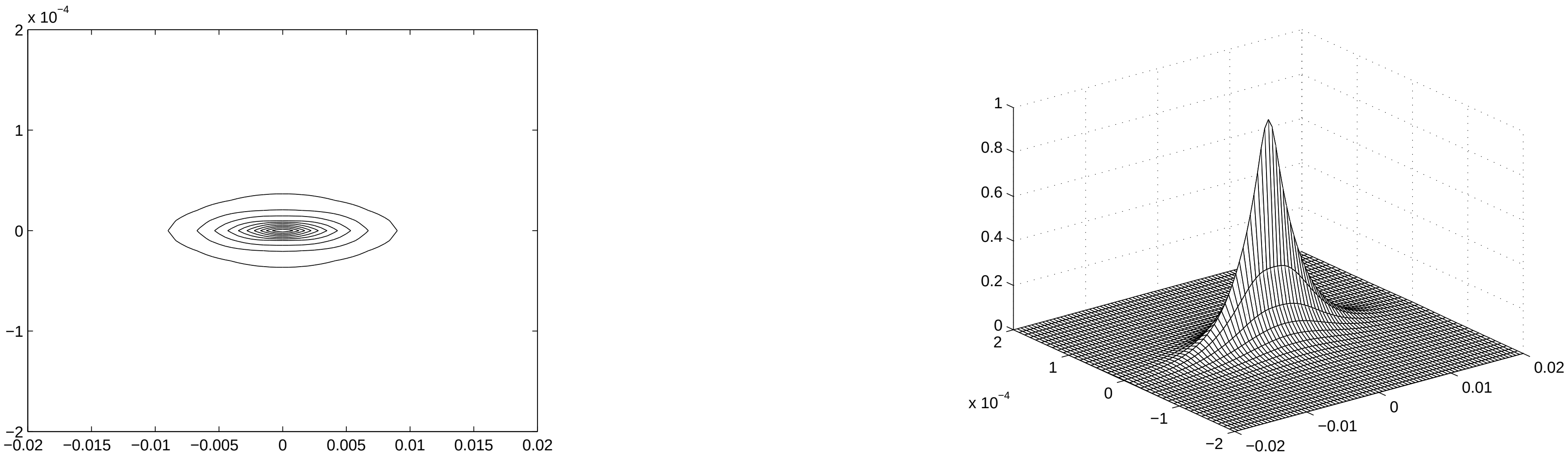


Figure 5: Contour and surface plots of the (scaled) profile likelihood for simulated data $(T = 1000)$ with $(\theta_1^{\text{true}}, \theta_2^{\text{true}}) = (0, 0)$

Choosing shrinkage priors for process variances

Take-away message from these figures:

- The information is contained in the likelihood function
- Don't destroy it by using the wrong prior
- For each process variance, a shrinkage prior should put positive prior probability to shrinking neighborhoods of 0 (Castillo and van der Vaart, 2012)

The non-centered state space model

The TVP model with initial value $\beta_{0,j} \sim \mathcal{N}\left(\beta_j, \theta_j P_{0j}\right)$,

$$\beta_{tj} = \beta_{t-1,j} + w_{jt}, \qquad w_{jt} \sim \mathcal{N}\left(0, \theta_j\right), \quad j = 1, \ldots, r,$$
$$y_t = \mathbf{x}_t \boldsymbol{\beta}_t + \varepsilon_t,$$

could be reparameterized as a **non-centered state space model**:

$$\tilde{\beta}_{jt} = \tilde{\beta}_{j,t-1} + \tilde{\omega}_{jt}, \quad \tilde{\omega}_{jt} \sim \mathcal{N}\left(0, 1\right), \quad j = 1, \ldots, r,$$
$$y_t = \mathbf{x}_t \boldsymbol{\beta} + \mathbf{x}_t \mathrm{Diag}\left(\sqrt{\theta_1}, \ldots, \sqrt{\theta_d}\right) \tilde{\boldsymbol{\beta}}_t + \varepsilon_t, \qquad (1)$$

with $\tilde{\beta}_{j0} \sim \mathcal{N}\left(0, P_{0j}\right)$ (Frühwirth-Schnatter and Wagner, 2010).

Dirac-spike-slab-priors for TVP models

Preferred choice: **Dirac-spike-and-slab prior** for β_j and $\sqrt{\theta_j}$:

$$\beta_j \sim (1-\omega^\delta)\Delta_0(\beta_j) + \omega^\delta \mathcal{N}\left(0, \tau_j^2\right),$$
$$\sqrt{\theta_j} \sim (1-\omega^\gamma)\Delta_0(\sqrt{\theta_j}) + \omega^\gamma \mathcal{N}\left(0, \xi_j^2\right),$$

i.e.:

- $\Pr(\beta_j = 0) = 1 - \omega^\delta$;
- $\Pr(\theta_j = 0) = 1 - \omega^\gamma$.

Choosing $\tau_j^2 \sim \mathcal{G}^{-1}(\nu_1, \nu_2)$ and $\xi_j^2 \sim \mathcal{G}^{-1}(\nu_3, \nu_4)$ leads to Student-t distributions in the slab (Castillo and van der Vaart, 2012).

Dirac-spike-slab-priors for TVP models

Representation involving binary indicators δ_j and γ_j with $\Pr(\delta_j = 1) = \omega^\delta$ and $\Pr(\gamma_j = 1) = \omega^\gamma$:

$$\beta_j|(\delta_j = 0) \equiv 0, \quad \beta_j|(\delta_j = 1) \sim \mathcal{N}\left(0, \tau_j^2\right),$$
$$\sqrt{\theta}_j|(\gamma_j = 0) \equiv 0, \quad \sqrt{\theta}_j|(\gamma_j = 1) \sim \mathcal{N}\left(0, \xi_j^2\right).$$

Classification of the coefficients based on the posterior probabilities $\Pr(\delta_j = 0, \gamma_j = 0|\mathbf{y})$, $\Pr(\delta_i = 1, \gamma_j = 0|\mathbf{y})$, and $\Pr(\delta_i = 1, \gamma_j = 1|\mathbf{y})$.

Dirac-spike-slab-priors for TVP models

- Perform MCMC based search in the model space defined by all 2^{2r} combinations of δ_j and γ_j (model search conditional on the latent variables $\tilde{\beta}_{jt}$).
- Use the posterior draws of δ_j and γ_j for model specification search.
- Allows intrinsic classification of each parameter into the three categories (insignificant, significant but fixed, dynamic).

MCMC for the Dirac-spike-and-slab prior

(a) Joint sampling of the indicators $\boldsymbol{\delta}, \boldsymbol{\gamma}$ from $p(\boldsymbol{\delta}, \boldsymbol{\gamma}|\mathbf{x}, \mathbf{y})$ using the marginal likelihood (closed form expression) of regression model (1) conditional on the state vector $\mathbf{x} = (\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_T)$

(b) Joint sampling of $\boldsymbol{\beta}, \sqrt{\theta_1}, \ldots, \sqrt{\theta_d}, \sigma^2$ conditional on $\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_T$ (inverted Gamma for σ^2, multivariate normal conditional on σ^2).

(c) Sample the states $\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_T$ in the NC state space model jointly conditional on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$: FFBS or AWOL for the **dynamic coefficients**, if $\theta_j = 0$ then simply sampled from the prior.

Sampling the state process

- **FFBS** (Frühwirth-Schnatter, 1994; Carter and Kohn, 1994; De Jong and Shephard, 1995; Durbin and Koopman, 2002)

- **A**ll **W**ith**O**ut a **L**oop (Kastner and Frühwirth-Schnatter, 2014): $\mathbf{x} = vec(\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_T) \sim \mathcal{N}\left(\Omega^{-1}c, \Omega^{-1}\right)$ where Ω is a very sparse band precision matrix (Rue, 2001; McCausland et al., 2011)
  - **No need to invert Ω**; instead fast band back-substitution
  - Cholesky decomposition of $\Omega = LL'$; solve $c = Lb$ for b
  - draw Td univariate standard normals, i.e. $\boldsymbol{\varepsilon} \sim \mathcal{N}\left(0, I_{Td}\right)$
  - solve $L'\mathbf{x} = b + \boldsymbol{\varepsilon}$ for $\mathbf{x}$.

Closely related to Chan and Jeliazkov (2009) (information filtering)

Dirac-spike-and-slab prior in practice

- Dirac-spike-and-slab priors for SSM and TVP model have been applied successfully in various papers (Frühwirth-Schnatter and Wagner, 2010; Proietti and Grassi, 2011; Grassi and Proietti, 2014).

- d large: single move MH of the indicators δ_j, γ_j for $j = 1, \ldots, d$ by sampling from $p(\delta_j, \gamma_j | (\boldsymbol{\delta}, \boldsymbol{\gamma})_{-j}, \mathbf{x}, \mathbf{y})$ may lead to convergence problems.

Need for alternative shrinkage priors.

Hierarchical shrinkage priors

Well-known for regression models ...

- **Bayesian Lasso prior** (Park and Casella, 2008):

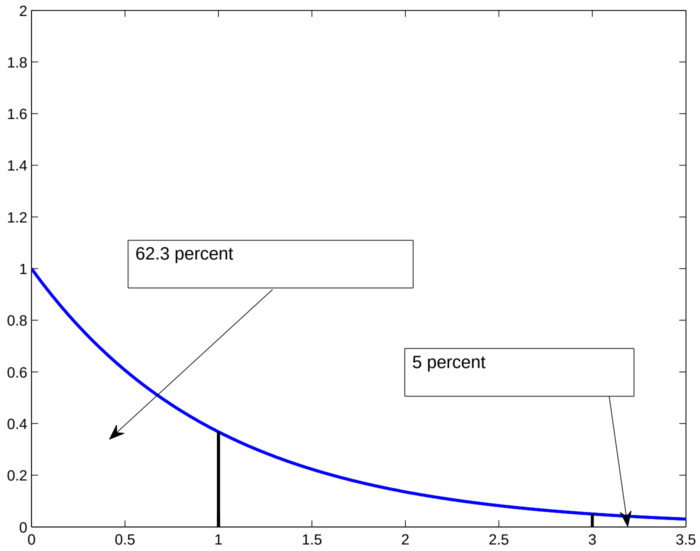
$$\beta_j|\psi_j \sim \mathcal{N}\left(0, 2/\lambda^2\psi_j\right), \qquad \psi_j \sim \mathcal{E}\left(1\right).$$

- **Normal-Gamma prior** (Griffin and Brown, 2010):

$$\beta_j|\psi_j \sim \mathcal{N}\left(0, 2/\lambda^2\psi_j\right), \qquad \psi_j \sim \mathcal{G}\left(a, a\right), \quad a < 1.$$

Shrink globally ($2/\lambda^2$), act locally ($\psi_j < 1$: more, $\psi_j > 1$: less shrinkage) (Polson and Scott, 2011)

Hierarchical shrinkage priors



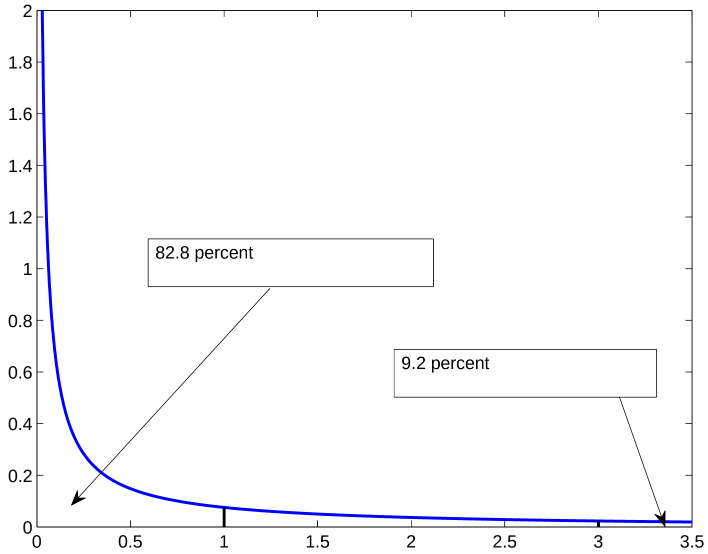


Figure 6: Shrinkage prior for ψ_i based on the exponential distribution (left) and the Gamma prior with $a = 0.1$ (right)

Hierarchical shrinkage priors for TVP models

Belmonte et al. (2014) use the **Bayesian Lasso** priors as shrinkage priors for the initial value β_j and the process variance $\sqrt{\theta_j}$.

Hierarchical representation:

$$\begin{aligned} \beta_j|\tau_j^2 &\sim \mathcal{N}\left(0, \tau_j^2\right), & \tau_j^2 &\sim \mathcal{E}\left(\lambda^2/2\right), \\ \sqrt{\theta_j}|\xi_j^2 &\sim \mathcal{N}\left(0, \xi_j^2\right), & \xi_j^2 &\sim \mathcal{E}\left(\kappa^2/2\right), \end{aligned} \tag{2}$$

A hierarchical prior is used for the level of overall shrinkage, i.e. $\lambda^2 \sim \mathcal{G}\left(\nu_1, \nu_2\right)$ and $\kappa^2 \sim \mathcal{G}\left(\nu_3, \nu_4\right)$.

The LASSO prior is known for overshrinking significant effects

More general regularization priors for TVP models

PhD project of Angela Bitto: more flexibility is obtained by using a **normal-Gamma prior** both for β_j and $\sqrt{\theta_j}$.
This prior has a hierarchical representation as in (2):

$$\begin{aligned} \beta_j|\tau_j^2 &\sim \mathcal{N}\left(0,\tau_j^2\right), \qquad & \tau_j^2 &\sim \mathcal{G}\left(a^\tau, a^\tau \lambda^2/2\right), \\ \sqrt{\theta_j}|\xi_j^2 &\sim \mathcal{N}\left(0,\xi_j^2\right), \qquad & \xi_j^2 &\sim \mathcal{G}\left(a^\xi, a^\xi \kappa^2/2\right). \end{aligned}$$

- Special case: $a^\tau = 1$ and $a^\xi = 1$ leads to Lasso hierarchical shrinkage prior.
- $a^\tau < 1$ and $a^\xi < 1$ introduces more sparsity without overshrinking.

MCMC for hierarchical shrinkage priors

(a) Sample the state $\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_T$ in the NC state space model jointly conditional on $\boldsymbol{\beta}$ and $\sqrt{\theta_1}, \ldots, \sqrt{\theta_d}$.

(b) Joint sampling of $(\boldsymbol{\beta}, \sqrt{\theta_1}, \ldots, \sqrt{\theta_d})$ **conditional on the prior variances τ_j^2 and ξ_j^2** and $\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_T$. The corresponding distribution is a multivariate normal distribution.

(c) Sample $\sigma^2|\cdot$ from an inverted Gamma distribution.

(d) Sample the prior variances τ_j^2 and ξ_j^2 conditional on $(\boldsymbol{\beta}, \sqrt{\theta_1}, \ldots, \sqrt{\theta_d})$ from independent **generalized inverse Gaussian (GIG) distributions**.

The generalized inverse Gaussian (GIG) distribution

The conditionally normal prior $\beta_j \sim \mathcal{N}\left(0, \tau_j^2\right)$ leads to a posterior for $\tau_j^2|\beta_j$, where the likelihood is the **kernel of an inverted Gamma density** in τ_j^2 which is combined with the **Gamma prior** $\tau_j^2 \sim \mathcal{G}\left(a^\tau, a^\tau \lambda^2/2\right)$:

$$\begin{aligned} p(\tau_j^2|\beta_j) &\propto (\tau_j^2)^{-1/2} \exp\left(-\frac{\beta_j^2}{2\tau_j^2}\right) \times (\tau_j^2)^{a^\tau - 1} \exp\left(-\frac{a^\tau \lambda^2 \tau_j^2}{2}\right) \\ &\propto (\tau_j^2)^{(a^\tau - 1/2) - 1} \exp\left(-\frac{a^\tau \lambda^2 \tau_j^2}{2}\right) \exp\left(-\frac{\beta_j^2}{2\tau_j^2}\right). \end{aligned}$$

The generalized inverse Gaussian (GIG) distribution

This leads to the **generalized inverse Gaussian (GIG) distribution** both for $\tau_j^2|\beta_j$ and $\xi_j^2|\theta_j$:

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

The inverse Gaussian distribution, $\mathcal{GIG}(p, a, b)$ is a three-parameter family with density ($a > 0$, $b > 0$ and p is a real parameter):

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

where $K_p(z)$ is the modified Bessel function of the second kind.

Sampling from the GIG distribution

A **very stable generator** is implemented in the recent R-package `GIGrvg` by J. Leydold (Hörmann and Leydold, 2013), see http://cran.r-project.org/web/packages/GIGrvg/index.html:

- Devroye (2013) shows that $Z = \log X$, where $X \sim \mathcal{GIG}\left(p, \sqrt{ab}, \sqrt{ab}\right)$ with $p > 0$, has a log-concave density.

- If $p > 0$, then $Y = \sqrt{b/a}e^{Z}$ follows $\mathcal{GIG}\left(p, a, b\right)$; if $p < 0$, then $Y = \sqrt{b/a}e^{-Z}$ follows $\mathcal{GIG}\left(-p, a, b\right)$.

MATLAB (`randraw`) or older R-packages fail with sampling from $\mathcal{GIG}\left(p, a, b\right)$, if b is close to 0 (natural under shrinkage priors!)

EU-aera inflation modelling

TVP version of the generalized Phillips curve (Belmonte et al., 2014):

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

where inflation depends on

- lags of inflation ($p = 12$) and 11 monthly seasonal dummies
- 13 other predictors $\mathbf{x}_t$ (1-month and 1-year Euribor, change in industrial production index, change in loans, change in monetary aggregate M3, unemployment rate, oil price, . . .)

EU-aera inflation modelling

- 37 possibly time-varying coefficients, each following $\beta_{tj} = \beta_{t-1,j} + w_{jt}, w_{jt} \sim \mathcal{N}(0, \theta_j)$ with unknown initial value β_j and unknown variance θ_j.

- Monthly data from February 1994 to November 2010

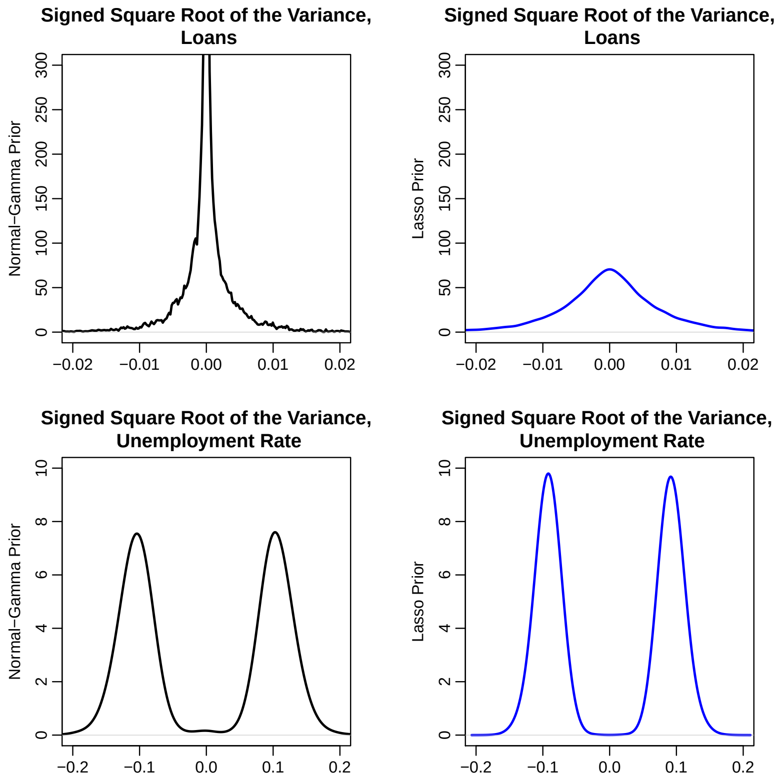
- LASSO priors (vary $\nu_1 = \nu_2 = \nu_3 = \nu_4 = \nu$)

- Normal-gamma priors (vary $a^\tau < 1$ and $a^\xi < 1$, vary $\nu_1 = \nu_2 = \nu_3 = \nu_4 = \nu$)

- Dirac-spike-and-slab priors lead to different solutions, depending on the starting values.

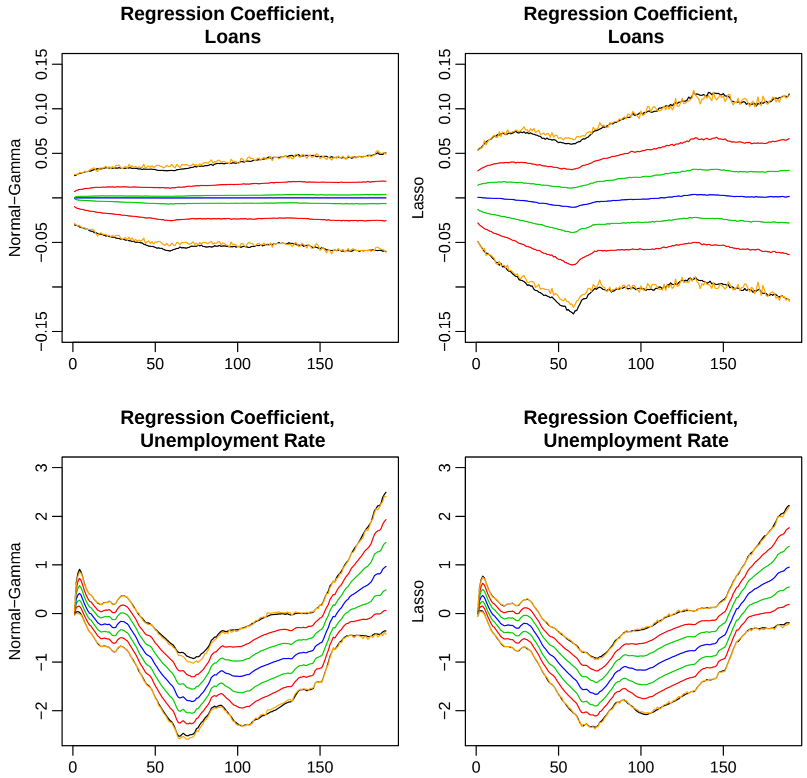
Posterior exploration for hierarchical shrinkage priors

Visual inspection of the posterior densities of $\sqrt{\theta_j}$ allows „classification" between static/dynamic:



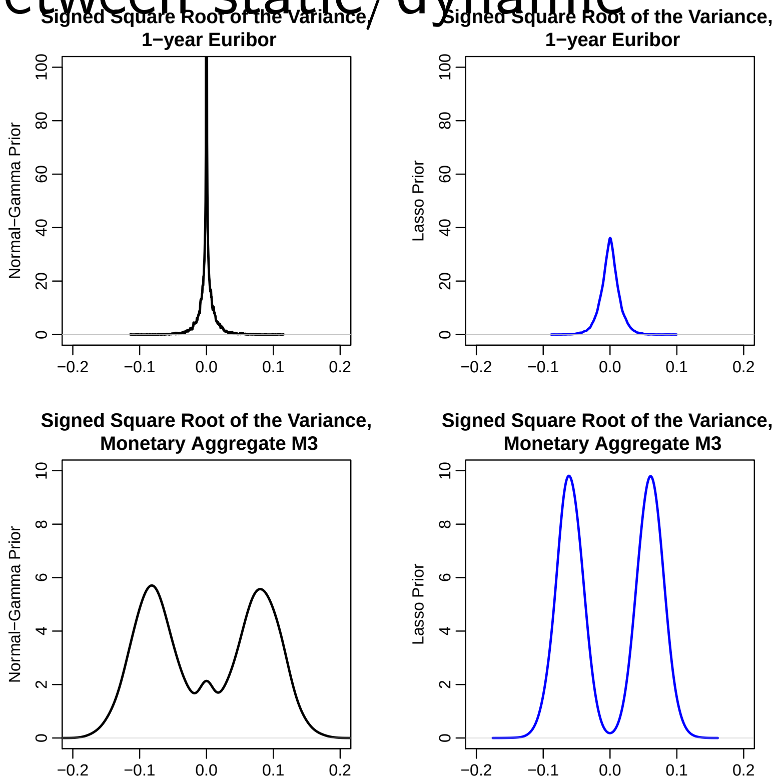
Posterior exploration for hierarchical shrinkage priors

Visual inspection of the posterior paths of $\beta_{tj}, t = 1, \ldots, 190$ (20,000 draws)



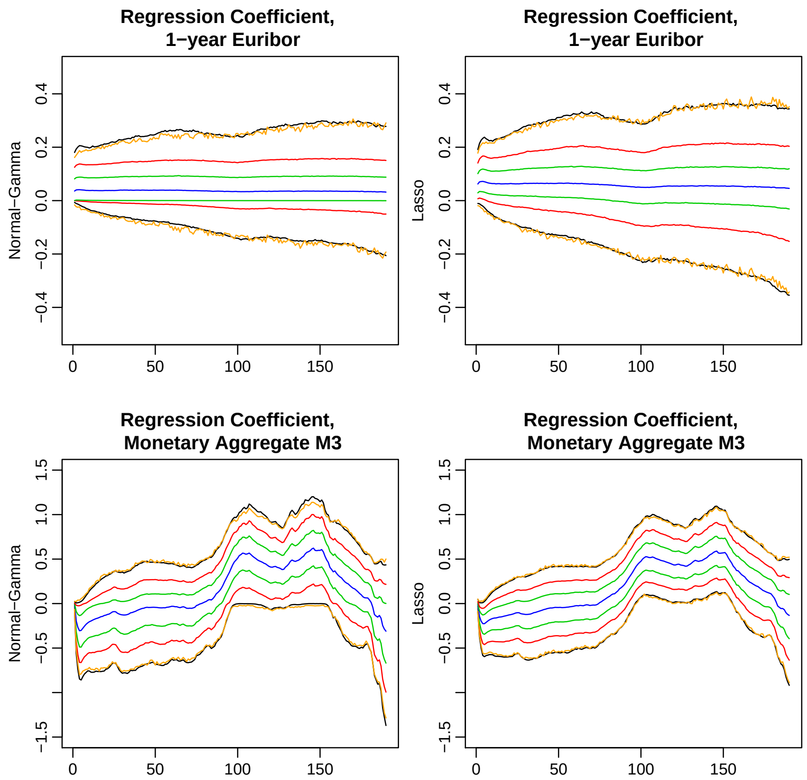
Posterior exploration for hierarchical shrinkage priors

Visual inspection of the posterior densities of $\sqrt{\theta_j}$ allows ,,classification" between static/dynamic



Posterior exploration for hierarchical shrinkage priors

Visual inspection of the posterior paths of $\beta_{tj}, t = 1, \ldots, 190$ (20,000 draws)



Predictive evaluation

Predictive evaluation for various shrinkage priors $\mathcal{M}$ over the last 60 months (i.e. $t_0 = 134$):

$$\log p(y_{t_0+1}, \ldots, y_T | \mathbf{y}^{\text{tr}}, \mathcal{M}) = \sum_{t=t_0+1}^{T} \log p(y_t | y_1, \ldots, y_{t-1}, \mathcal{M}).$$

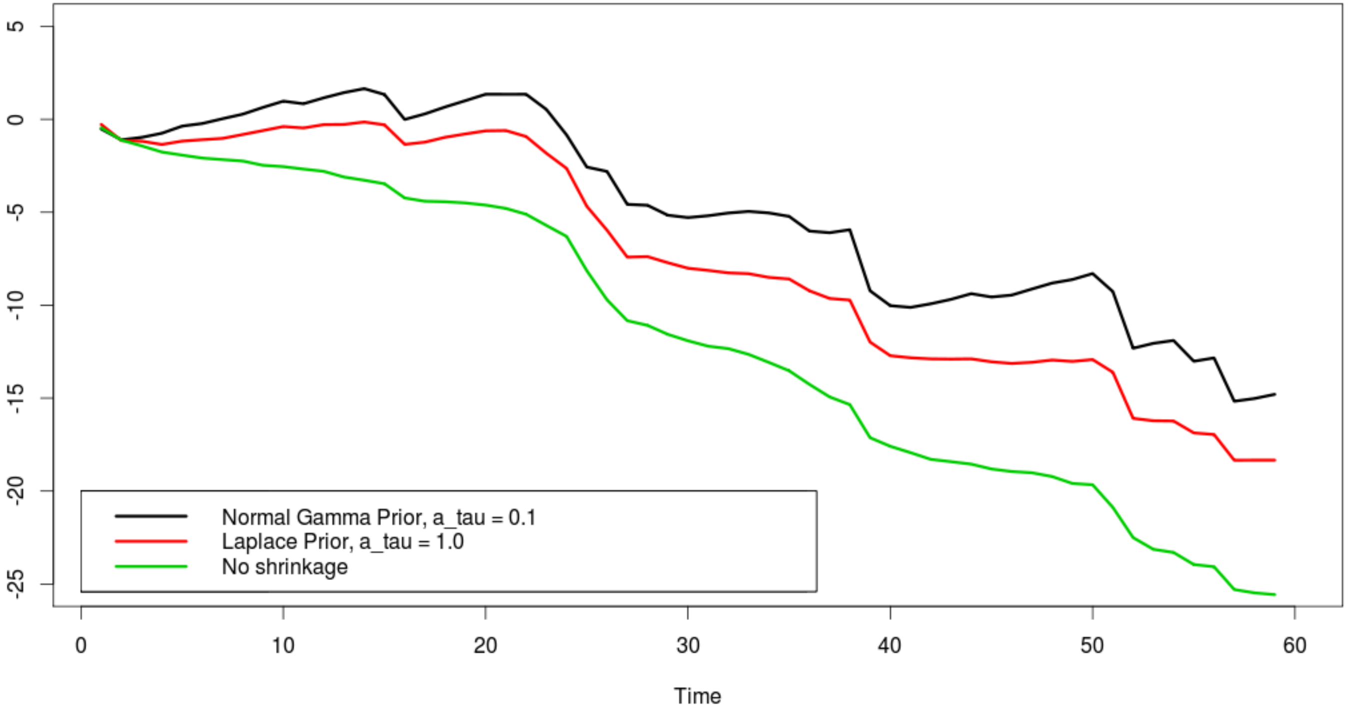
- Interpretation as training sample marginal likelihood for shrinkage prior $\mathcal{M}$.
- For each $t = t_0 + 1, \ldots, T$ independent MCMC runs on local cluster (500 cores) to determine $p(y_t | y_1, \ldots, y_{t-1}, \mathcal{M})$ for each $\mathcal{M}$.

Predictive evaluation for EU-aera inflation modelling

What is a good hierarchical shrinkage prior?

- First results suggest that **spike-slab priors (if they converge)** beat hierarchical shrinkage priors
- The normal-Gamma with small $a^{\tau} < 1$ and $a^{\xi} < 1$ beats the LASSO prior considered in Belmonte et al. (2014)
- All priors dominate models without shrinkage.

Predictive evaluation for EU-aera inflation modelling



Extensions of the standard TVP model

- **Stochastic volatility** using the R-package `stochvol` (Kastner and Frühwirth-Schnatter, 2014)

- Extend $\mathbf{Q} = \mathrm{Var}(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1})$ to a **full covariance matrix** as for covariance selection in random effects models (Chen and Dunson, 2003; Frühwirth-Schnatter and Tüchler, 2008)

- **Non-Gaussian** TVP models using auxiliary mixture sampling (Frühwirth-Schnatter and Wagner, 2010)

- **Multivariate** TVP models (financial application)

Cholesky SV Modelling - Multivariate TVP models

A multivariate time series $\mathbf{y}_t \sim \mathcal{N}_r(\mathbf{0}, \mathbf{\Sigma}_t)$ with time-varying covariance matrix may be represented as a set of time-varying regressions (Lopes et al., 2012), e.g. for $r = 3$:

$$
\begin{aligned}
y_{1t} &= \varepsilon_{1t}, \qquad \varepsilon_{1t} \sim \mathcal{N}\left(0, \sigma^2_{1,t}\right) \\
y_{2t} &= \beta_{21t} y_{1t} + \varepsilon_{2t}, \qquad \varepsilon_{2t} \sim \mathcal{N}\left(0, \sigma^2_{2,t}\right) \\
y_{3t} &= \beta_{31t} y_{1t} + \beta_{32t} y_{2t} + \varepsilon_{3t}, \qquad \varepsilon_{3t} \sim \mathcal{N}\left(0, \sigma^2_{3,t}\right)
\end{aligned}
$$

with state equation:

$$
\beta_{ij,t} = \beta_{ij,t-1} + w_{ijt}, \qquad w_{ijt} \sim \mathcal{N}\left(0, \theta_{ij}\right).
$$

More details on Cholesky SV modelling

Lopes et al. (2012) consider the Cholesky decomposition of $\mathbf{\Sigma}_t$:

$$\mathbf{\Sigma}_t = \boldsymbol{A}_t \boldsymbol{D}_t \boldsymbol{A}_t^{'},$$

where $\boldsymbol{A}_t$ is a lower triangular (LT) with ones on the main diagonal and $\boldsymbol{D}_t$ is a diagonal matrix. Then

$$\boldsymbol{A}_t^{-1} \mathbf{y}_t \sim \mathcal{N}_r \left(\mathbf{0}, \boldsymbol{A}_t^{-1} \mathbf{\Sigma}_t (\boldsymbol{A}_t^{'})^{-1} \right) = \mathcal{N}_r \left(\mathbf{0}, \boldsymbol{D}_t \right). \tag{3}$$

Since $\boldsymbol{A}_t^{-1}$ is also a LT with ones on the main diagonal, (3) may be rewritten with $\mathbf{B}_t = -(\boldsymbol{A}_t^{-1} - \mathbf{I}_r)$ is LT with zero main diagonal:

$$\mathbf{y}_t = \mathbf{B}_t \mathbf{y}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_r \left(\mathbf{0}, \boldsymbol{D}_t \right).$$

Application to DAX data

Daily stock returns indices from the DAX from September 4th 2001 to August 31st, 2011

$T = 2500$, $r = 29$ series

- Filtered using univariate SV-models to reduce conditional heteroscedasticity (Frühwirth-Schnatter and Lopes, 2012)
- Unfiltered returns using the TVP-SV model

Model search with SSVS

3 (filtered) stock returns from the DAX: BMW/Daimler/Volkswagen (Car industry)

Indicators highest probability model (frequency equals 0.54):

$\beta_{21,0}$	$\beta_{31,0}$	$\beta_{32,0}$
0	**1**	**1**
θ_{21}	θ_{31}	θ_{32}
0	0	**1**

Indicators median probability model (Scott and Berger, 2006):

$\beta_{21,0}$	$\beta_{31,0}$	$\beta_{32,0}$
0	**1.0**	**1.0**
θ_{21}	θ_{31}	θ_{32}
0.	0.45	**0.99**

Posterior exploration

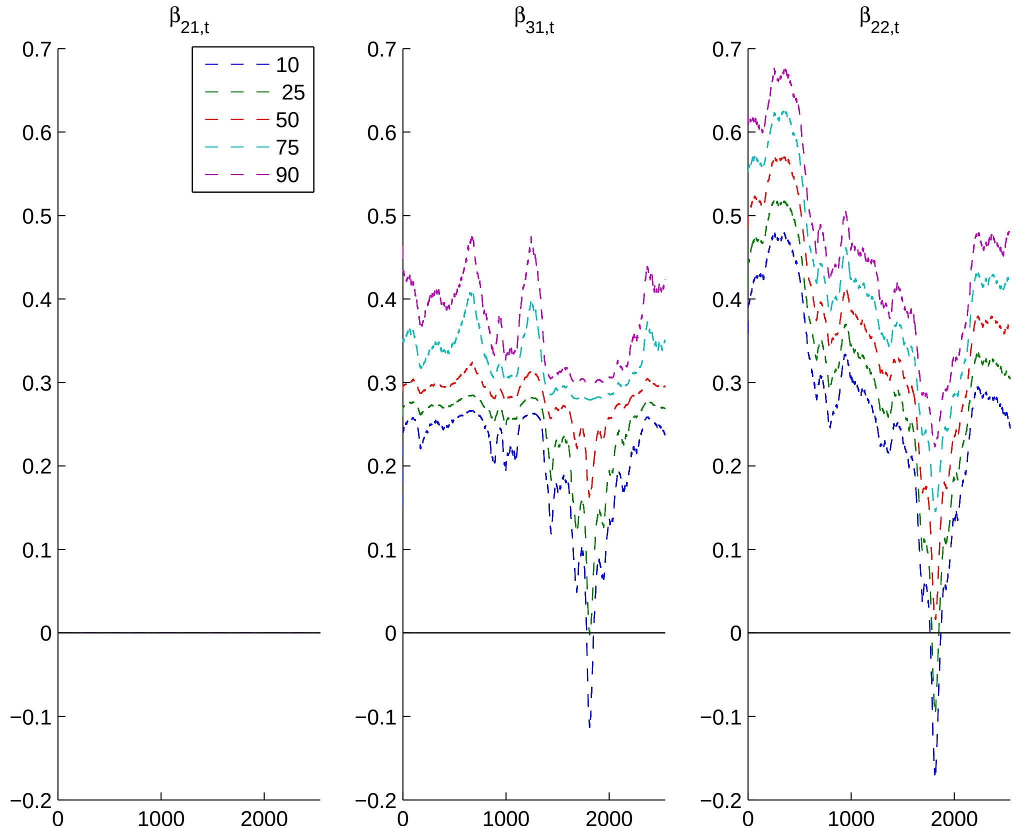


Figure 7: Posterior simulations for $\beta_{t,21}$, $\beta_{t,31}$ and $\beta_{t,32}$

Posterior exploration

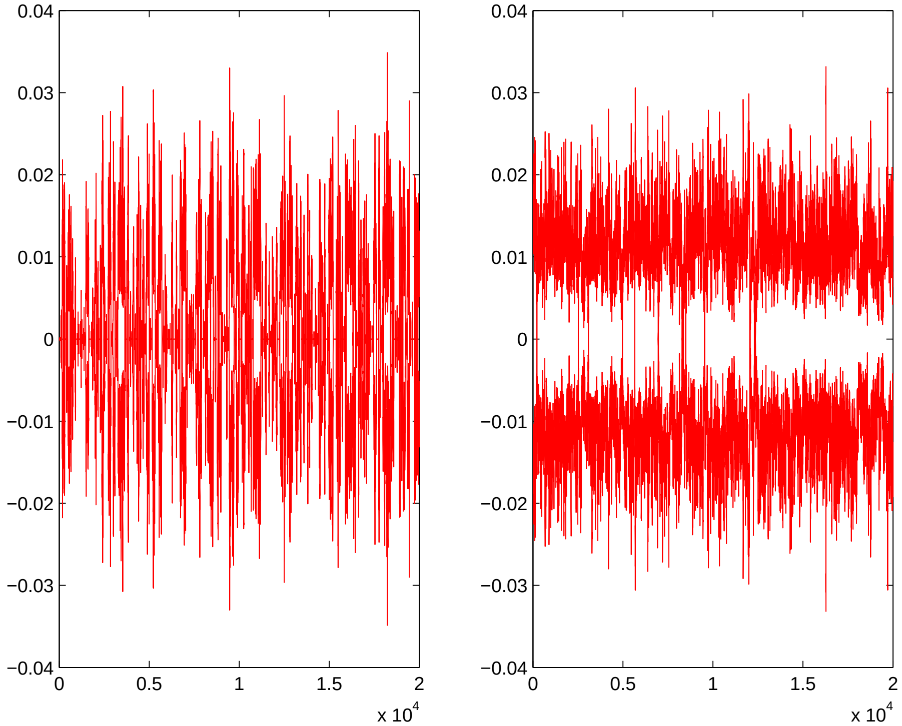


Figure 8: Posterior simulations for $\sqrt{\theta_{t,31}}$ and $\sqrt{\theta_{t,32}}$

Posterior exploration

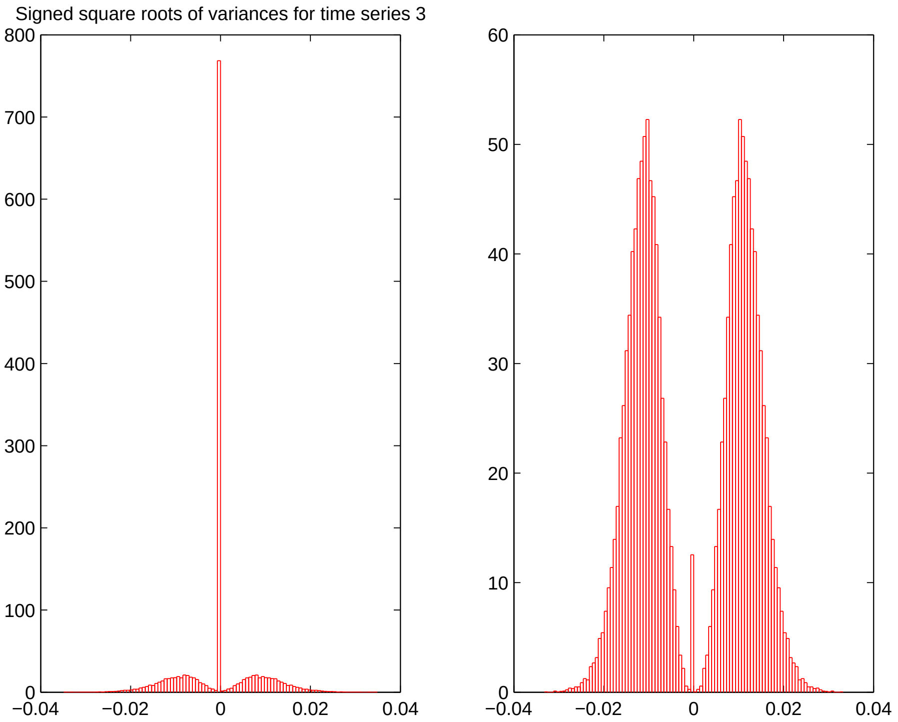


Figure 9: Posterior distributions for $\sqrt{\theta_{t,31}}$ and $\sqrt{\theta_{t,32}}$

Global minimum variance portfolio

Time-varying weights w_{t1}, w_{t2}, and w_{t3} for the global minimum variance portfolio $P_t = \mathbf{w}_t'\mathbf{y}_t$

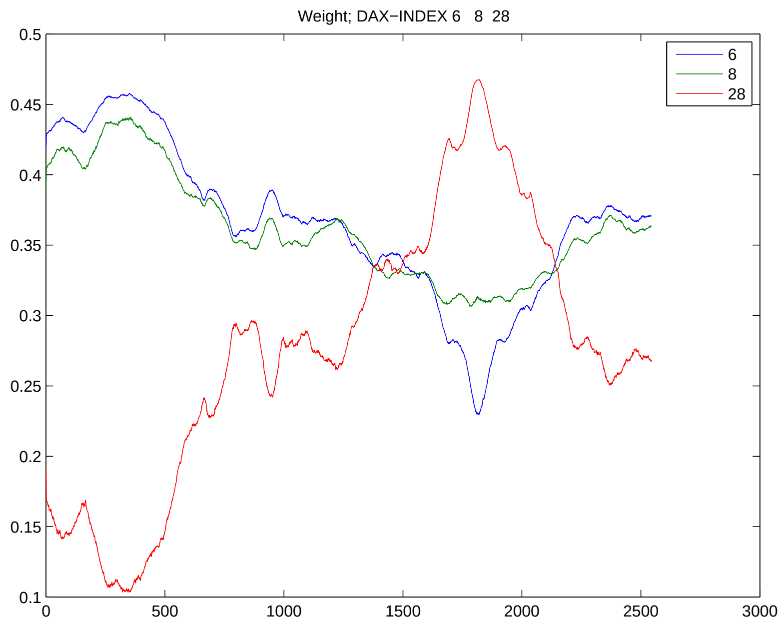


Figure 10: 6 BMW; 8 Daimler; 28 Volkswagen

What difference does it make?

Fully dynamic model (green) versus parsimonious model using spike-slab-priors (red)

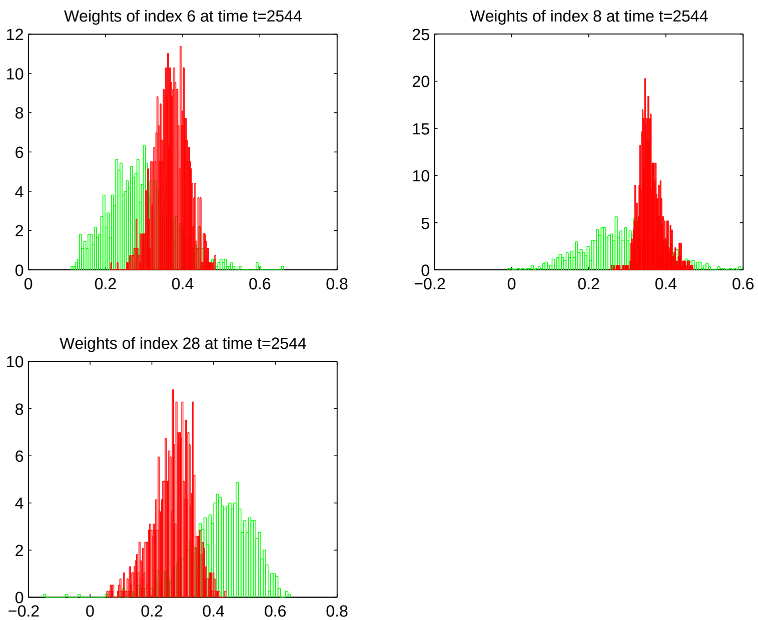


Figure 11: Weights w_{t1}, w_{t2}, and w_{t3} for the global minimum variance portfolio $P_t = \mathbf{w}_t'\mathbf{y}_t$ at $t = T$

All 29 (filtered) DAX returns

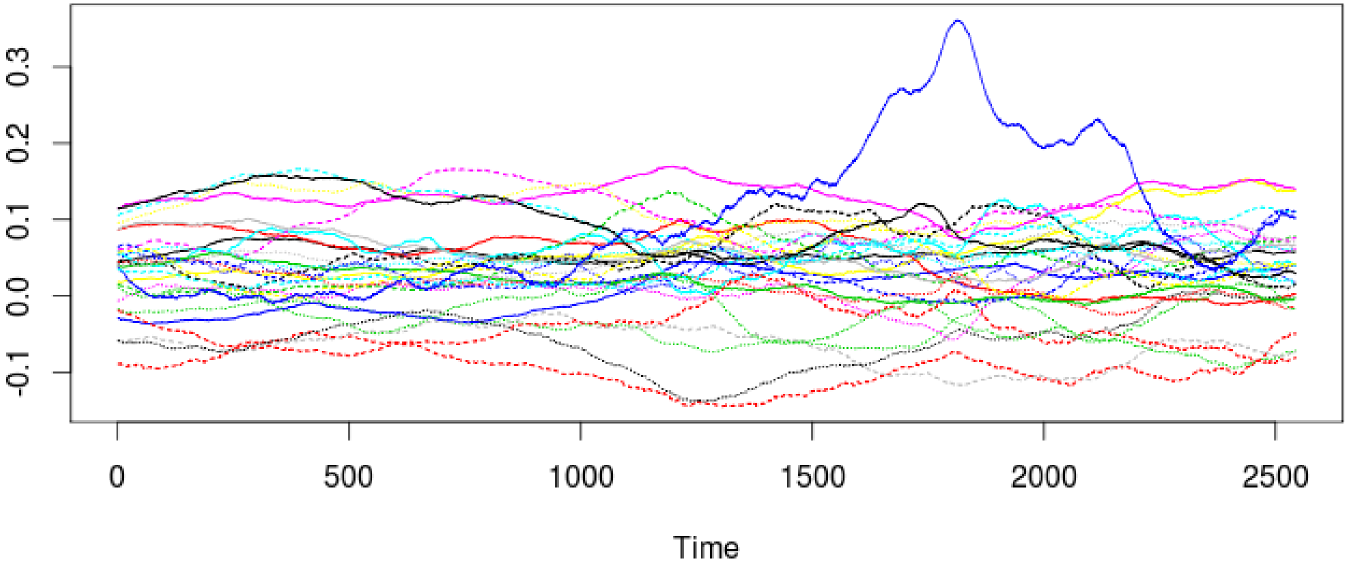
- Leads to $29 \cdot 28/4 = 406$ (!) dynamic coefficients $\beta_{ij,t}$

- Variable selection among $3^{406} = 5 \cdot 10^{193}$ models $\Rightarrow$ great challenge for the Dirac-spike-and-slab prior

- LASSO priors (vary $\nu_1 = \nu_2 = \nu_3 = \nu_4 = \nu$)

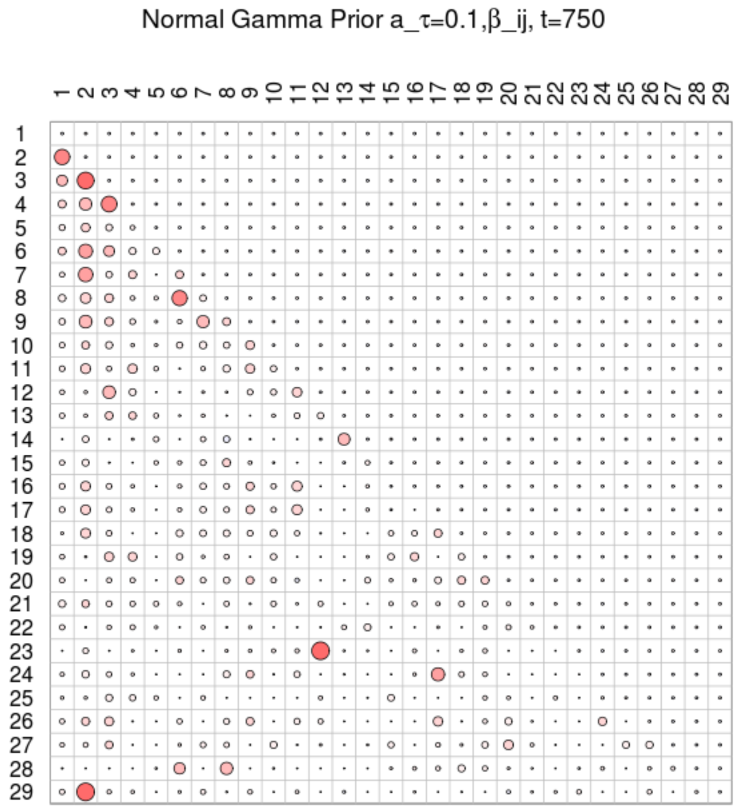
- Normal-gamma priors (vary $a^\tau < 1$ and $a^\xi < 1$, vary $\nu_1 = \nu_2 = \nu_3 = \nu_4 = \nu$)

All 29 (filtered) DAX returns



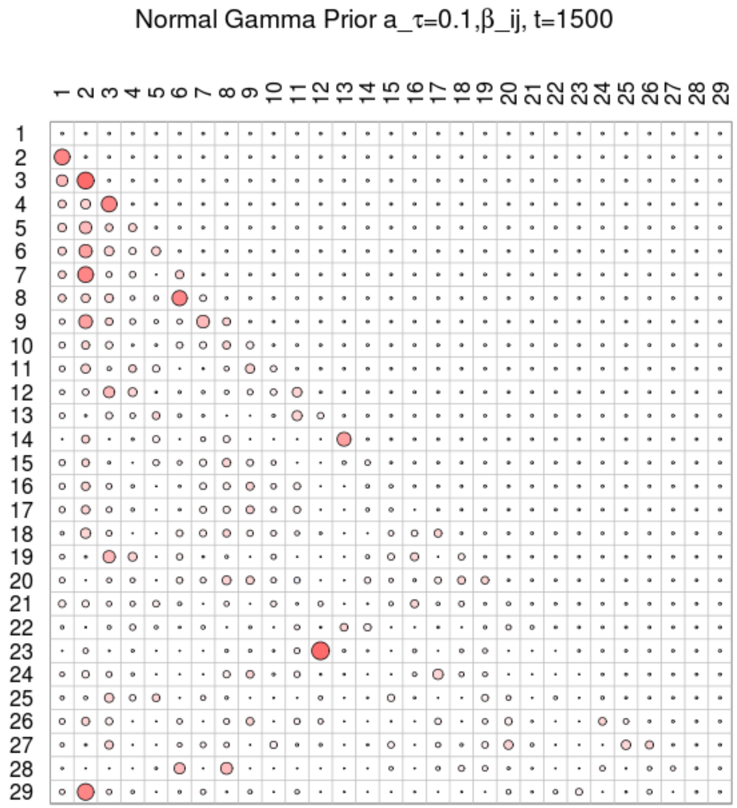
Time-varying portfolio weights (normal Gamma Prior, $a_\tau = 0.1$)

Sparsity



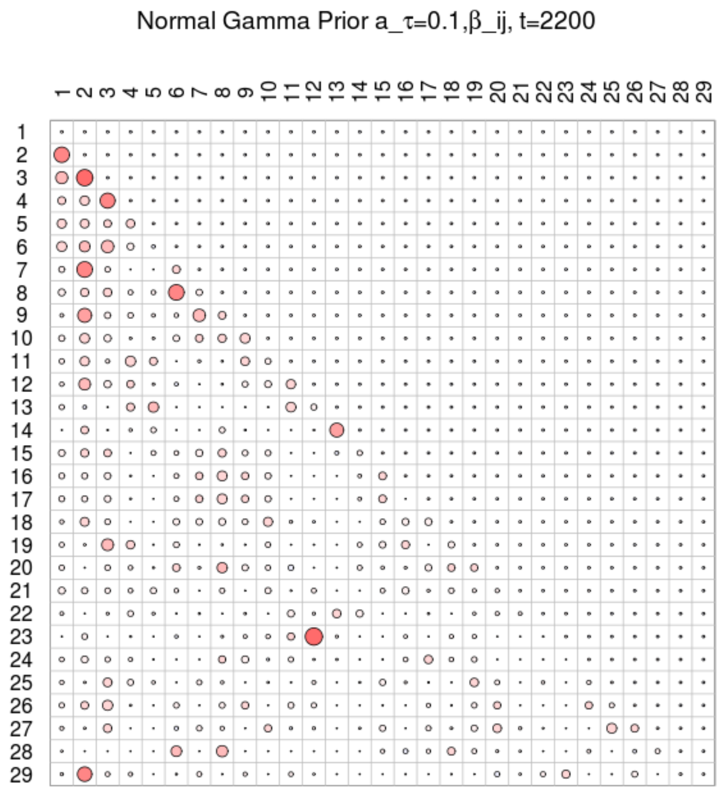
Cholesky factors $\beta_{ij,t}$ at $t = 750$

Sparsity



Cholesky factors $\beta_{ij,t}$ at $t = 1500$

Sparsity



Cholesky factors $\beta_{ij,t}$ at $t = 2200$

Summary

- Well-known ideas of shrinkage and variable selection for the standard regression model can be easily extended to TVP models using a non-centered version of the state space model

- Hierarchical representation allows efficient MCMC sampling

- Normal gamma hierarchical shrinkage priors seem preferable to the LASSO; strongly outperforms standard TVP without shrinkage

- Spike-and-slab priors allow model search/variable selection, but are unstable for higher dimensional models

Outlook

- More theoretical work in the spirit of Castillo and van der Vaart (2012) needed for TVP models (model consistency, oracle property)
- Incorporate recent work on time-varying sparsity (Nakajima and West, 2013; Kalli and Griffin, 2014)
- Apply to (high-dimensional) data sets from

  Economics (VAR models with TVP coefficients)
  Finance (returns and other financial time series)
  and
  Business (marketing research)

References

Belmonte, M. A. G., Koop, G., and Korobolis, D. (2014). Hierarchical shrinkage in time-varying parameter models. Technical report, Journal of Forecasting, forthcoming.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. Biometrika, 81:541–553.

Castillo, I. and van der Vaart, A. (2012). Needles and straws in a haystack: Posterior concentration for possibly sparse sequences. The Annals of Statistics, 40:2069–2101.

Chan, J. C. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. International Journal of Mathematical Modelling and Numerical Optimisation, 1:101–120.

References

Chen, Z. and Dunson, D. (2003). Random effects selection in linear mixed models. Biometrics, 59:762–769.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. Journal of Financial Economics, XX:forthcoming.

De Jong, P. and Shephard, N. (1995). The simulation smoother for time series models. Biometrika, 82:339–350.

Devroye, L. (2013). Random variate generation for the generalized inverse Gaussian distribution. Statistics and Computing, 23:XX–XX.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis.

References

Biometrika, 89:603–615.

Eisenstat, E., Chan, J. C., and Strachan, R. W. (2014). Stochastic model specification search for time-varying parameter VARs. XX, XX:XX–XX.

Fahrmeir, L., Kneib, T., and Konrath, S. (2010). Bayesian regularisation in structured additive regression: A unifying perspective on shrinkage, smoothing and predictor selection. Statistics and Computing, 20:203–219.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. Journal of Time Series Analysis, 15:183–202.

Frühwirth-Schnatter, S. (2004). Efficient Bayesian parameter estimation. In Harvey, A., Koopman, S. J., and Shephard, N.,

References

editors, State Space and Unobserved Component Models: Theory and Applications, pages 123–151. Cambridge University Press, Cambridge.

Frühwirth-Schnatter, S. and Lopes, H. (2012). Parsimonious Bayesian Factor Analysis when the Number of Factors is Unknown. Research report, Booth School of Business, University of Chicago.

Frühwirth-Schnatter, S. and Tüchler, R. (2008). Bayesian parsimonious covariance estimation for hierarchical linear mixed models. Statistics and Computing, 18:1–13.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for Gaussian and partially non-Gaussian state space models. Journal of Econometrics, 154:85–100.

References

Frühwirth-Schnatter, S. and Wagner, H. (2011). Bayesian variable selection for random intercept modeling of Gaussian and non-Gaussian data. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, Bayesian Statistics 9, pages 165–200. Oxford University Press, Oxford (UK).

Grassi, S. and Proietti, T. (2014). Characterizing economic trends by bayesian stochastic model specification search. Computational Statistics and Data Analysis, missing:XX–XX.

Griffin, J. E. and Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. Bayesian Analysis, 5:171–188.

References

Hörmann, W. and Leydold, J. (2013). Generating generalized inverse gaussian random variates. Statistics and Computing, 23:forthcoming.

Kalli, M. and Griffin, J. E. (2014). Time-varying sparsity in dynamic regression models. forthcoming:XX–XX.

Kastner, G. and Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models. Computational Statistics and Data Analysis, page forthcoming.

Lopes, H. F., McCulloch, R. E., and Tsay, R. S. (2012). Cholesky stochastic volatility models for high-dimensional time series. Discussion Paper.

References

McCausland, W. J., Miller, S., and Pelletier, D. (2011). Simulation smoothing for state space models: A computational efficiency analysis. Computational Statistics and Data Analysis, 55:199 – 212.

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

Park, T. and Casella, G. (2008). The Bayesian Lasso. Journal of the American Statistical Association, 103:681–686.

Polson, N. G. and Scott, J. G. (2011). Shrink globally, act locally: Sparse Bayesian regularization and prediction. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D.,

References

Smith, A. F. M., and West, M., editors, Bayesian Statistics 9, pages 501–538. Oxford University Press, Oxford (UK).

Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. Review of Economic Studies, 72:821–852.

Proietti, T. and Grassi, S. (2011). Bayesian stochastic model specification search for seasonal and calendar effects. In Bell, W. R., Holan, S. H., and McElroy, T. S., editors, Economic Time Series: Modelling and Seasonality, chapter 1. ??, ?? Preliminary electronic version, not printed.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. Journal of the Royal Statistical Society, Ser. B, 63:325–338.

Scott, J. G. and Berger, J. O. (2006). An exploration of aspects

References

of Bayesian multiple testing. Journal of Statistical Planning and Inference, 136:2144–2162.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Ser. B, 58:267–288.

Wagner, H. and Duller, C. (2012). Bayesian model selection for logistic regression models with random intercept. Computational Statistics and Data Analysis, 56:1256–1274.