Stochastic Model Specification Search for Gaussian and Non-Gaussian State Space Models

Sylvia Frühwirth-Schnatter and Helga Wagner

Johannes Kepler Universität Linz

Non-regular Model Selection Problems

Model selection for state space models often leads to testing problems which are non-regular from the view-point of classical statistics

- Observation equation for time series observation y_t, $t = 1, \ldots, T$:

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2_\varepsilon\right)$$

- μ_t follows a random walk with a random drift:

$$\begin{aligned} \mu_t &= \mu_{t-1} + a_{t-1} + \omega_{1t}, \qquad \omega_{1t} \sim \mathcal{N}\left(0, \theta_1\right) \\ a_t &= a_{t-1} + \omega_{2t}, \qquad \omega_{2t} \sim \mathcal{N}\left(0, \theta_2\right), \end{aligned}$$

Non-regular Model Selection Problems

- Is the drift a_t fixed or random? Testing $\theta_2 = 0$ versus $\theta_2 > 0$: the null hypothesis lies on the boundary of the parameter space
- Is the drift term a_t significant? Testing the null hypothesis $a_0 = a_1 = \cdots = a_T = 0$ (local level model) versus a local trend model is a non-regular problem: size of the hypothesis increases with the number of observations

No standard likelihood-ratio tests.

The Bayesian Approach

The Bayesian approach is, in principle, able to deal with non-regular testing problems

Assign a prior probability $p(\mathcal{M}_k)$ to each model (e.g. uniform distribution over all models)

Compute the posterior probability distribution $p(\mathcal{M}_k|\mathbf{y})$ for each model using Bayes rule

$$p(\mathcal{M}_k|\mathbf{y}) \propto p(\mathbf{y}|\mathcal{M}_k)p(\mathcal{M}_k),$$

where $p(\mathbf{y}|\mathcal{M}_k)$ is the marginal likelihood for model $\mathcal{M}_k$.

Challenges with the Bayesian Approach

Posterior probabilities are not easily computed:

- Marginal Likelihoods: Computing the posterior probabilities $p(\mathcal{M}_k|\mathbf{y})$ by Bayes rule requires computation of the marginal likelihood $p(\mathbf{y}|\mathcal{M}_k)$ for each model. As the integration may be high-dimensional this is a numerical challenge: importance sampling (Zellner and Rossi, 1984; Frühwirth-Schnatter, 1995), Chib's estimator (Chib, 1995), bridge sampling (Frühwirth-Schnatter, 2004), auxiliary mixture sampling (Frühwirth-Schnatter and Wagner, 2008a)

- Model-space MCMC methods – sample jointly the model indicator and the unknown parameters
  - reversible jump MCMC (Green, 1995)
  - variable selection approach (George and McCulloch, 1993, 1997)

The Variable Selection Approach for Regression Models

For regression models the variable selection approach aims at identifying non-zero regression effects by indicators:

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \quad i = 1, \ldots, N$$

$$\begin{array}{ll} \beta_j = 0, & \text{iff } \delta_j = 0, \\ \beta_j \text{ unconstrained}, & \text{iff } \delta_j = 1. \end{array}$$

Each restricted regression model may be represented by a certain realization of $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_d)'$, e.g. $d = 5$:

$$\begin{aligned} y_i &= x_{i1}\beta_1 + x_{i4}\beta_4 + \beta_5 + \varepsilon_i \\ &= \delta_1 x_{i1}\beta_1 + \delta_2 x_{i2}\beta_2 + \delta_3 x_{i3}\beta_3 + \delta_4 x_{i4}\beta_4 + \delta_5\beta_5 + \varepsilon_i, \\ & \boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5)' = (1, 0, 0, 1, 1)'. \end{aligned}$$

The Variable Selection Approach for Regression Models

- Joint estimation of $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_d)$, $\boldsymbol{\beta}$ and σ^2_ε by a simple MCMC scheme

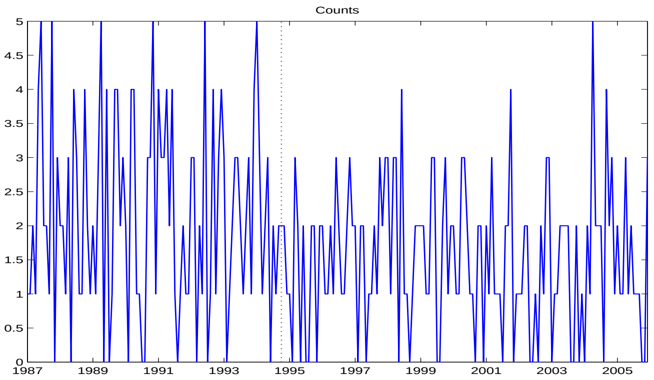
- Feasible, even if the total number of possible models is rather large; attractive alternative to computing marginal likelihoods also for a small set of models

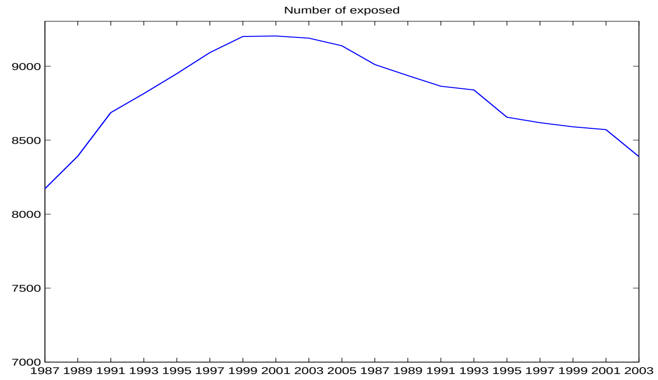
- Feasible for non-Gaussian data (Tüchler, 2008) using auxiliary mixture sampling (Frühwirth-Schnatter and Wagner, 2006; Frühwirth-Schnatter and Frühwirth, 2007)

- It is useful far beyond the common problem of selecting covariates and allows
  - covariance selection in random effects models (Chen and Dunson, 2003; Frühwirth-Schnatter and Tüchler, 2008).
  - Model selection for state space models (Frühwirth-Schnatter and Wagner, 2008b)

Time Series of Road Accidents of Children

killed or injured pedestrians, Children (aged 6-10) in Linz (Austria) (left) and number of children in this age (right), monthly data January 1987 - December 2005





A legal intervention intended to increase road safety became effective on **October 1,1994**: increased priority for pedestrians

Time Series of Road Accidents of Children

Basic structural model (Harvey and Durbin, 1986):

$$
\begin{aligned}
&y_t \sim \mathcal{P}\left(e_t \lambda_t\right),\\
&\log \lambda_t = \mu_t + s_t,\\
&\mu_t = \mu_{t-1} + a_{t-1} + \omega_{1t}, \qquad \omega_{1t} \sim \mathcal{N}\left(0, \theta_1\right)\\
&a_t = a_{t-1} + \omega_{2t}, \qquad \omega_{2t} \sim \mathcal{N}\left(0, \theta_2\right),\\
&s_t = -s_{t-1} - \cdots - s_{t-11} + \omega_{3t}, \qquad \omega_{3t} \sim \mathcal{N}\left(0, \theta_3\right)
\end{aligned}
$$

Modification of trend component for time of intervention, $t = t_{int}$:

$$
\mu_t = \mu_{t-1} + a_{t-1} + \Delta + \omega_{1t}
$$

Time Series of Road Accidents of Children

Unrestricted model: intervention effect not significant!?

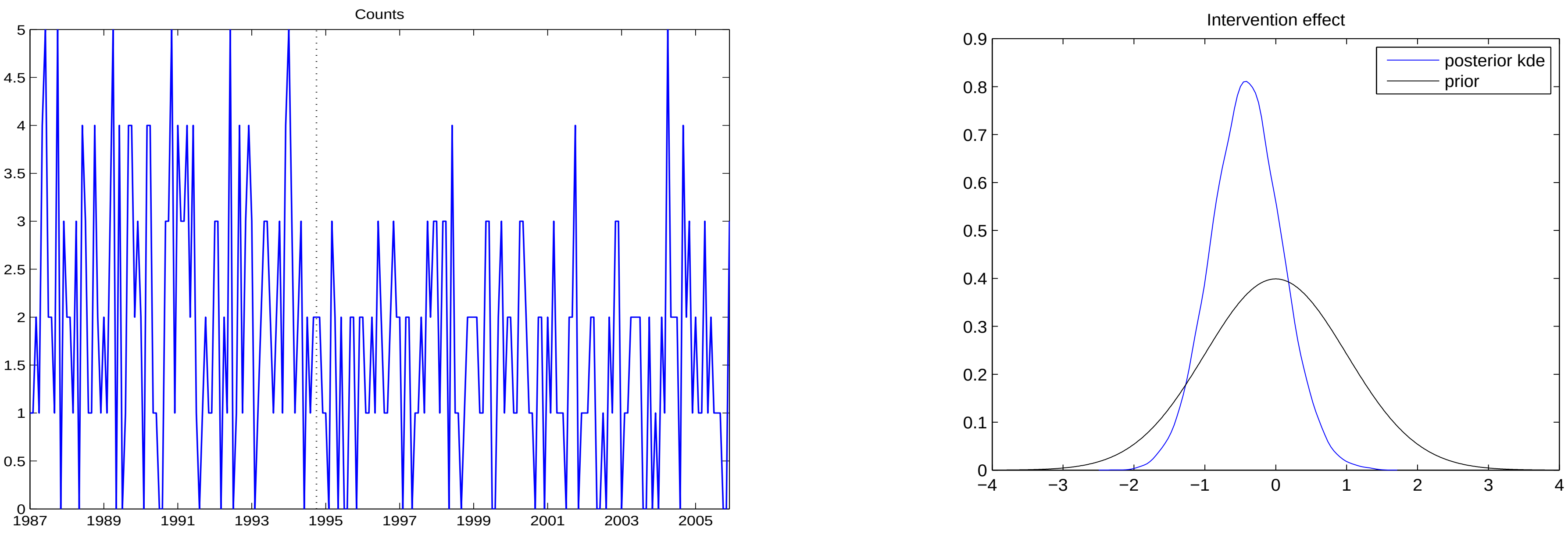


Figure 1: Posterior density of the intervention effect Δ (October 1,1994) in comparison to the prior (unrestricted basic structural model)

Stochastic model specification search for state space models

- How to introduce binary indicators for model selection?
- How to choose the priors?
- How to run MCMC?

Discuss Frühwirth-Schnatter and Wagner (2008b)

The Parsimonious Local Trend Model

Introduce three binary indicator δ, γ_1 and γ_2 in a noncentred version of the model:

$$\tilde{\mu}_t = \tilde{\mu}_{t-1} + \tilde{\omega}_{1t}, \qquad \tilde{\omega}_{1t} \sim \mathcal{N}\left(0, 1\right), \tag{1}$$

$$\tilde{a}_t = \tilde{a}_{t-1} + \tilde{\omega}_{2t}, \qquad \tilde{\omega}_{2t} \sim \mathcal{N}\left(0, 1\right), \tag{2}$$

$$\tilde{A}_t = \tilde{A}_{t-1} + \tilde{a}_{t-1}, \tag{3}$$

$$y_t = \mu_0 + \delta t a_0 + \gamma_1 \sqrt{\theta_1} \tilde{\mu}_t + \gamma_2 \sqrt{\theta_2} \tilde{A}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2_\varepsilon\right), \tag{4}$$

with $\tilde{\mu}_0 = \tilde{a}_0 = \tilde{A}_0 = 0$. μ_0 and a_0 are the initial values for the level and the drift component and θ_1 and θ_2 are equal to the variances in the dynamic linear trend model.

The Parsimonious Local Trend Model

8 different models:

- $\delta = 1, \gamma_1 = 1, \gamma_2 = 1$ leads to the local trend model
- $\delta = 0, \gamma_2 = 0, \gamma_1 = 1$ leads to the local level model
- $\gamma_1 = 0, \gamma_2 = 0, \delta = 1$ leads to a regression model with linear trend
- 5 additional models

Identifiability

The sign of $\sqrt{\theta_j}$ and $\tilde{\mu}_t, \tilde{A}_t$ is not identified, because it may be changed without changing the likelihood function, e.g.:

$$y_t \quad = \quad \ldots + \sqrt{\theta_1}\tilde{\mu}_t + \ldots + \varepsilon_t = \ldots + (-\sqrt{\theta_1})(-\tilde{\mu}_t) + \ldots + \varepsilon_t.$$

Similarly, the sign of $\sqrt{\theta_2}$ and the sequences $\{\tilde{a}_t\}_1^T$ and $\{\tilde{A}_t\}_1^T$ may be changed without changing the distribution of $y_1, \ldots, y_T$.

To make this unidentifiability transparent we write the observation equation as:

$$y_t \quad = \quad \mu_0 + \delta_1 t a_0 + \gamma_1(\pm\sqrt{\theta_1})\tilde{\mu}_t + \gamma_2(\pm\sqrt{\theta_2})\tilde{A}_t + \varepsilon_t. \tag{5}$$

All 4 parameters $\boldsymbol{\vartheta} = (\pm\sqrt{\theta_1}, \pm\sqrt{\theta_2}, \sigma^2_\varepsilon, \mu_0, a_0)$ define the same integrated likelihood. As a consequence, the likelihood function $p(\mathbf{y}|\boldsymbol{\vartheta})$ is symmetric around 0 in the direction of $\sqrt{\theta_1}$ and $\sqrt{\theta_2}$ and therefore multimodal.

Identifiability

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

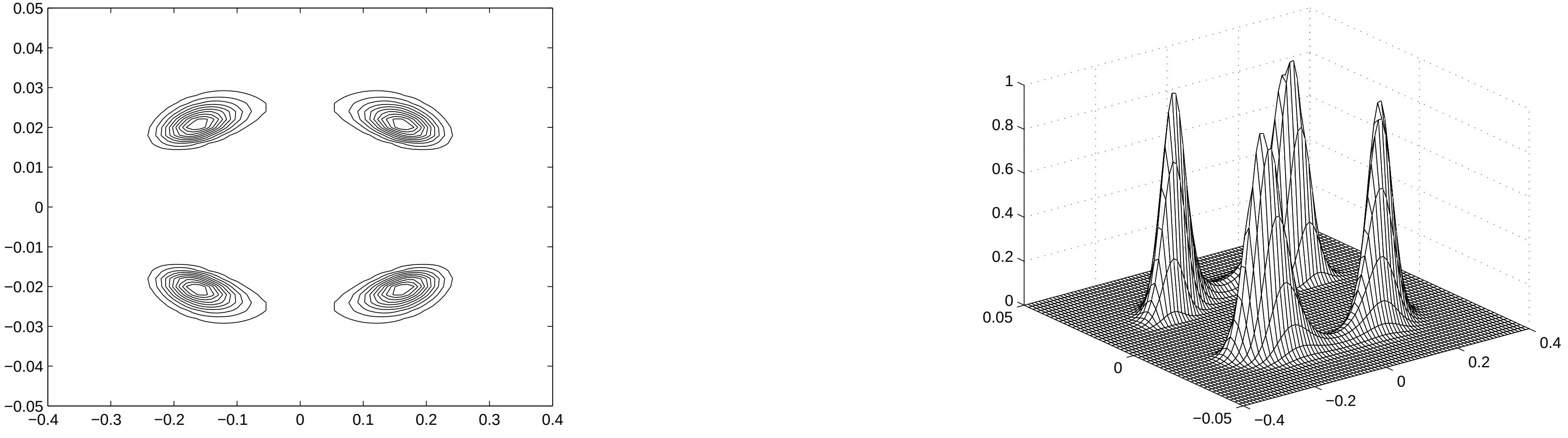
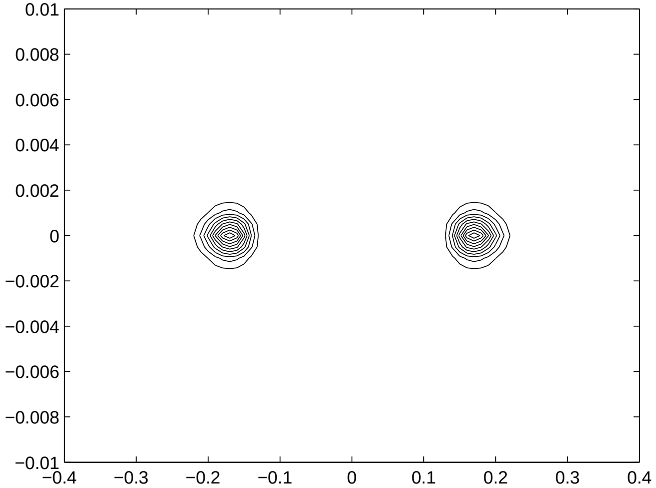


Figure 2: Contour and surface plots of the (scaled) profile likelihood $l(\sqrt{\theta_1}, \sqrt{\theta_2})/\max(l(\sqrt{\theta_1}, \sqrt{\theta_2}))$, where $l(\sqrt{\theta_1}, \sqrt{\theta_2}) = p(\mathbf{y}|\sqrt{\theta_1}, \sqrt{\theta_2}, \sigma_\varepsilon^{2,\text{true}}, \mu_0^{\text{true}},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

Identifiability

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.



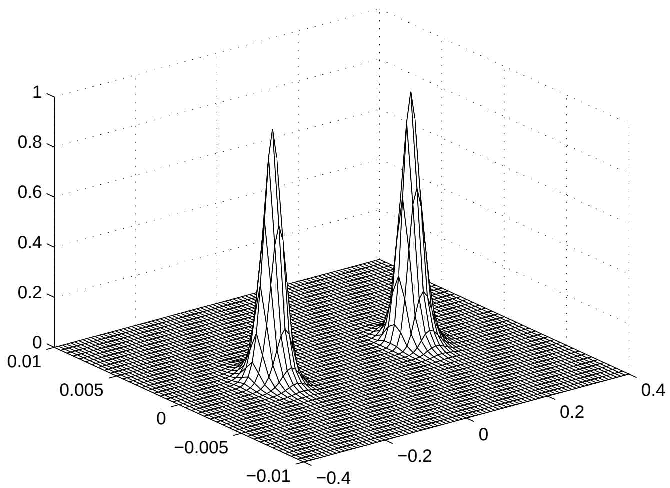


Figure 3: Contour and surface plots of the (scaled) profile likelihood for simulated data $(T = 1000)$ with $(\theta_1^{\text{true}}, \theta_2^{\text{true}}) = (0.15, 0)$

Identifiability

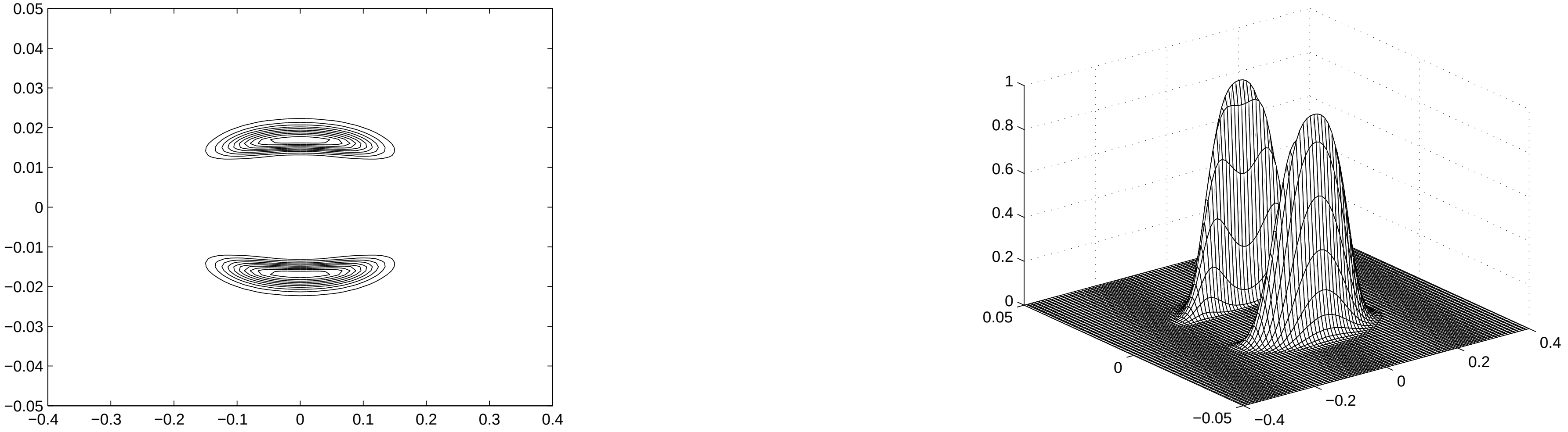


Figure 4: Contour and surface plots of the (scaled) profile likelihood for simulated data ($T = 1000$) with $(\theta_1^{\mathrm{true}}, \theta_2^{\mathrm{true}}) = (0, 0.02)$

Identifiability

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.

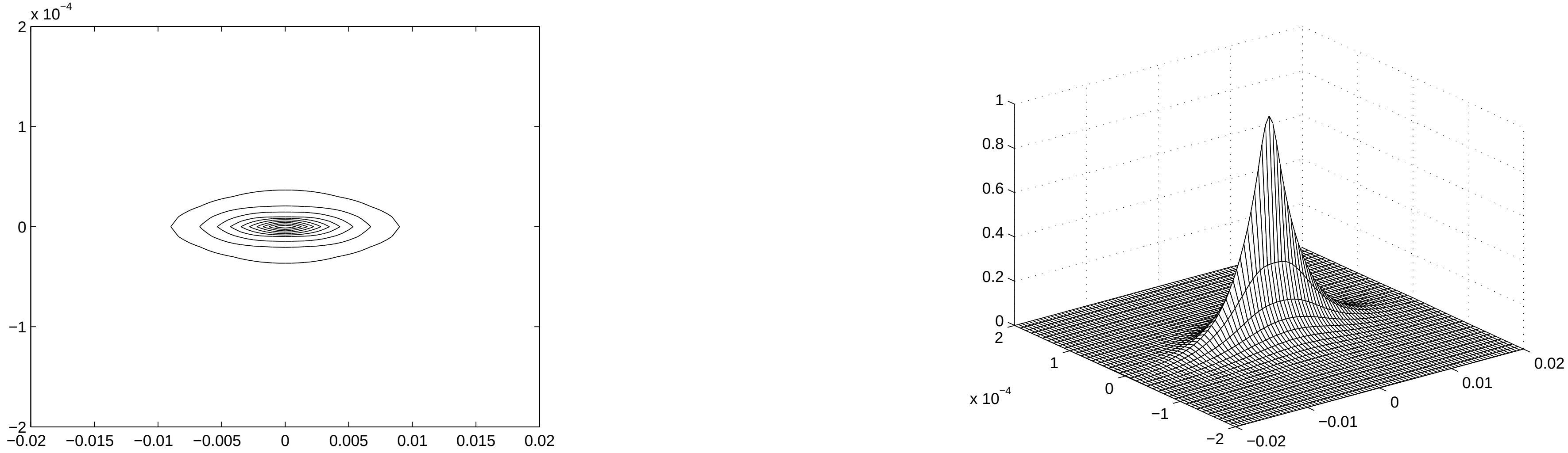


Figure 5: Contour and surface plots of the (scaled) profile likelihood for simulated data ($T = 1000$) with $(\theta_1^{\text{true}}, \theta_2^{\text{true}}) = (0, 0)$

The non-centered parameterization

Observation equation of the non-centered parameterization:

$$y_t = \mu_0 + \delta_1 t a_0 + \gamma_1(\pm\sqrt{\theta_1})\tilde{\mu}_t + \gamma_2(\pm\sqrt{\theta_2})\tilde{A}_t + \varepsilon_t. \tag{6}$$

- Extension to other state space models: introduce separate indicators for the fixed and the really dynamic part

- Natural conjugate prior for $\boldsymbol{\beta} = (\mu_0, a_0, \pm\sqrt{\theta_1}, \pm\sqrt{\theta_2})$ in (4) is a normal distribution

- Conditional on the state vector $(\tilde{\mu}_t, \tilde{A}_t)$, standard variable selection problem in regression model (4)

- State equation independent of model parameters, full conditional Gibbs sampling more efficient than for the usual (centered) parameterization

Extension to the basic structural model

Initial (unknown) seasonal pattern $\mathbf{s}_0 = (s_{-S}, \ldots, s_0)$ with $s_{-S} + \ldots + s_0 = 0$.

Additional binary indicators δ_2 and γ_3 for the fixed and the really dynamic part of the seasonal pattern:

- $\delta_2 = 1, \gamma_3 = 1$ time-varying seasonal pattern
- $\delta_2 = 1, \gamma_3 = 0$ fixed seasonal pattern
- $\delta_2 = 0, \gamma_3 = 0$ no seasonal pattern

Introduce the indicators into the non-centered parameterization

The non-centered parameterization

For the seasonal component, the non-centered parameterization is based on following stochastic difference equation:

$$\tilde{s}_t = -\tilde{s}_{t-1} - \cdots - \tilde{s}_{t-S+1} + \tilde{\omega}_{3t}, \qquad \tilde{\omega}_{3t} \sim \mathcal{N}\left(0, 1\right), \tag{7}$$

where $\tilde{s}_{-S+1} = \ldots = \tilde{s}_0 = 0$. Combine state equation (7) with the state equations (1) to (3) and following observation equation:

$$\begin{aligned} y_t \quad = \quad & \mu_0 + \delta_1 t a_0 + \delta_2 s_{j(t)} \\ & + \gamma_1 (\pm\sqrt{\theta_1}) \tilde{\mu}_t + \gamma_2 (\pm\sqrt{\theta_2}) \tilde{A}_t + \gamma_3 (\pm\sqrt{\theta_3}) \tilde{s}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2_\varepsilon\right). \end{aligned} \tag{8}$$

$s_{j(t)}$ is the initial seasonal component corresponding to time t.

Choosing the prior

Example: local level model

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right) \\
\mu_t &= \mu_{t-1} + \omega_{1t}, \qquad \omega_{1t} \sim \mathcal{N}\left(0, \theta\right).
\end{aligned}
$$

Prior on the variance θ may be influential when testing $\theta = 0$ versus $\theta > 0$, if the true value is close to 0. Compare:

- the standard conditional conjugate prior, $\theta \sim \mathcal{G}^{-1}\left(c_0, C_0\right)$
- a normal prior on the signed square root, $\pm\sqrt{\theta} \sim \mathcal{N}\left(0, B_0\right)$

$\pm\sqrt{\theta}$ is the signed square root of the variance

Choosing the prior

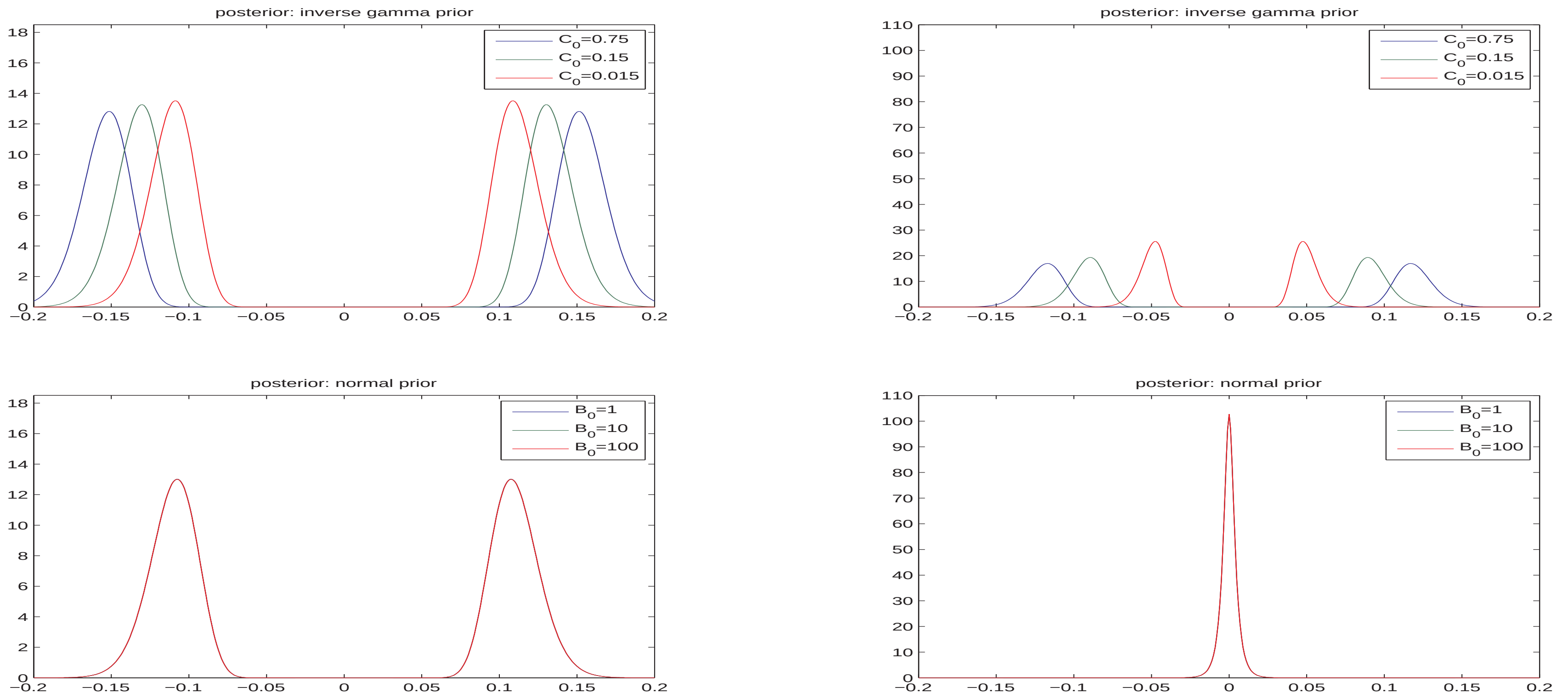


Figure 6: Posterior density for $\pm\sqrt{\theta}$ under different priors; top: $\theta \sim \mathcal{G}^{-1}(0.5, C_0)$, bottom: $\pm\sqrt{\theta} \sim \mathcal{N}(0, B_0)$; left: dynamic model with $\theta = 0.01$; right: static model with $\theta = 0$; $\sigma_\varepsilon^2 = 1$, $T = 100$

Choosing the prior for the basic structural model

Use a hierarchical prior for σ_ε^2: $\sigma_\varepsilon^2 \sim \mathcal{G}^{-1}(c_0, C_0)$ and $C_0 \sim \mathcal{G}(g_0, G_0)$.

Use normal priors for μ_0, a_0, and $\mathbf{s}_0$ as usual

We do not use the usual inverted Gamma prior for $\theta_1, \ldots, \theta_3$. The parameters $\pm\sqrt{\theta_1}, \pm\sqrt{\theta_2}$ and $\pm\sqrt{\theta_3}$ are coefficients in a regression model, use a normal prior:

- partially proper prior where $p(\mu_0) \propto \mathrm{constant}$, use $\mathcal{N}(0, B_0)$ for all unrestricted parameters

- fractional prior (O'Hagan, 1995) for all unrestricted parameters including μ_0

Assume a prior distribution for the indicators $\boldsymbol{\delta} = (\delta_1, \delta_2)$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)$ (e.g. uniform distribution).

MCMC estimation

Gibbs sampling scheme:

1. Sampling of the indicators $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ from $p(\boldsymbol{\delta}, \boldsymbol{\gamma}|\mathbf{x}, \mathbf{y})$ using the regression model (8) conditional on the state vector $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$, $\mathbf{x}_t = (\tilde{\mu}_t, \tilde{a}_t, \tilde{A}_t, \tilde{s}_t)$ (closed form expression because the model is conditionally normal given $\boldsymbol{\beta}$);

2. Conditional on $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$ and $\mathbf{x}$, sampling of the unrestricted initial values μ_0, a_0, $\mathbf{s}_0$ and the unrestricted variances $\pm\sqrt{\theta_j}$ using the regression model (6)

3. Sampling of $\mathbf{x}$ from the non-centered state space model using forward-filtering-backward-sampling (Frühwirth-Schnatter, 1994; Carter and Kohn, 1994; De Jong and Shephard, 1995; Durbin and Koopman, 2002)

4. Perform independent random sign switch for: $\pm\sqrt{\theta_1}$ and $\{\tilde{\mu}_t\}_{t=1}^T$; $\pm\sqrt{\theta_2}$ and $\{\tilde{a}_t, \tilde{A}_t\}_{t=1}^T$; $\pm\sqrt{\theta_3}$ and $\{\tilde{s}_t\}_{t=1}^T$

Example: UK coal consumption data

UK coal consumption Harvey (1989): quarterly data from 1/1960 to 4/1986

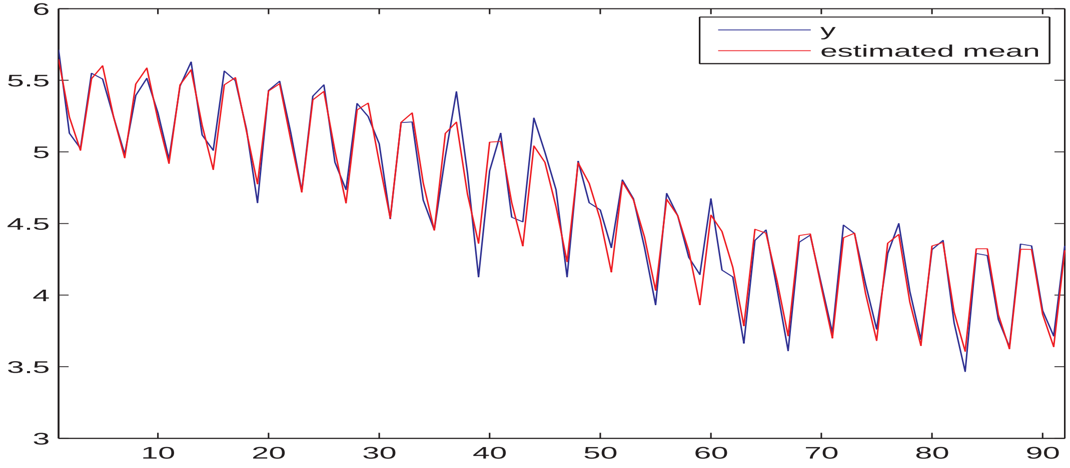


Figure 7: UK coal consumption 1960-1986 (log scale)

Comparing the Centered and the Non-centered Parameterization

Influence of the prior; no variable selection ($\boldsymbol{\delta} = 1$ and $\boldsymbol{\gamma} = 1$)

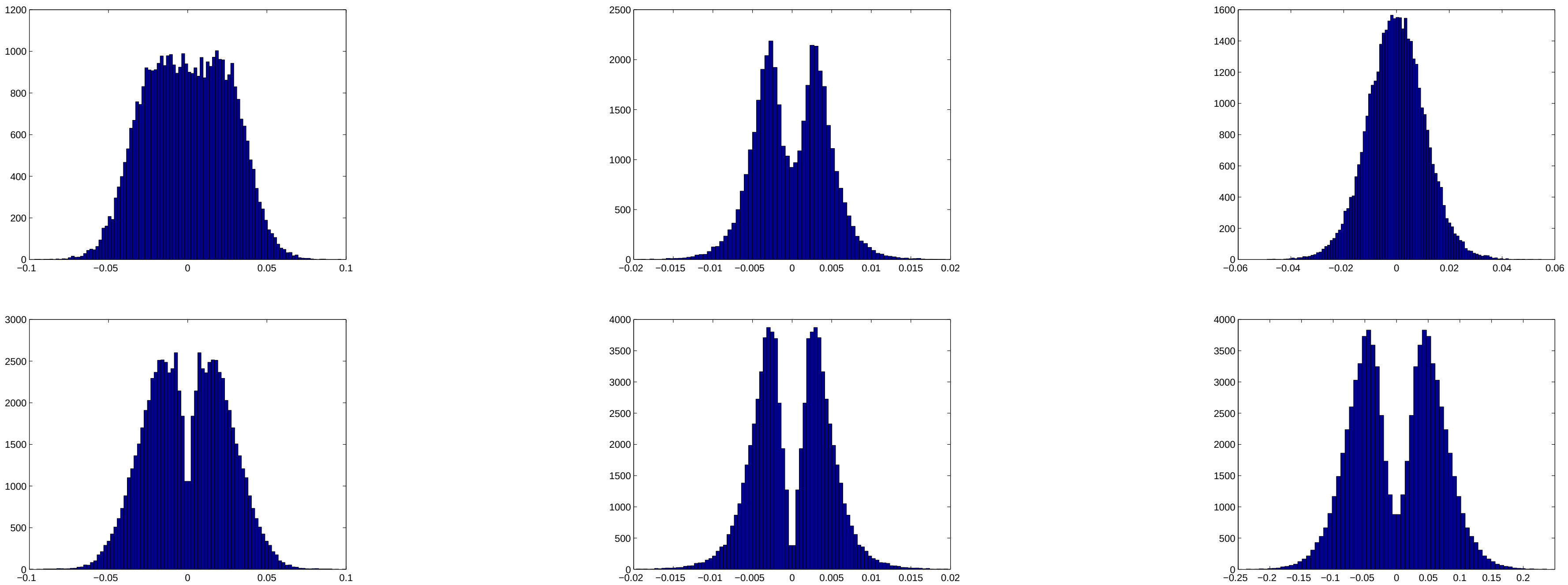


Figure 8: Histograms for $\pm\sqrt{\theta_1}$ (left), $\pm\sqrt{\theta_2}$ (middle) and $\pm\sqrt{\theta_3}$ (right); top: $\mathcal{N}(0,1)$ prior for $\pm\sqrt{\theta_i}$; bottom: $\mathcal{G}^{-1}\left(-0.5, 10^{-7}\right)$-prior for θ_i

Comparing the Centered and the Non-centered Parameterization

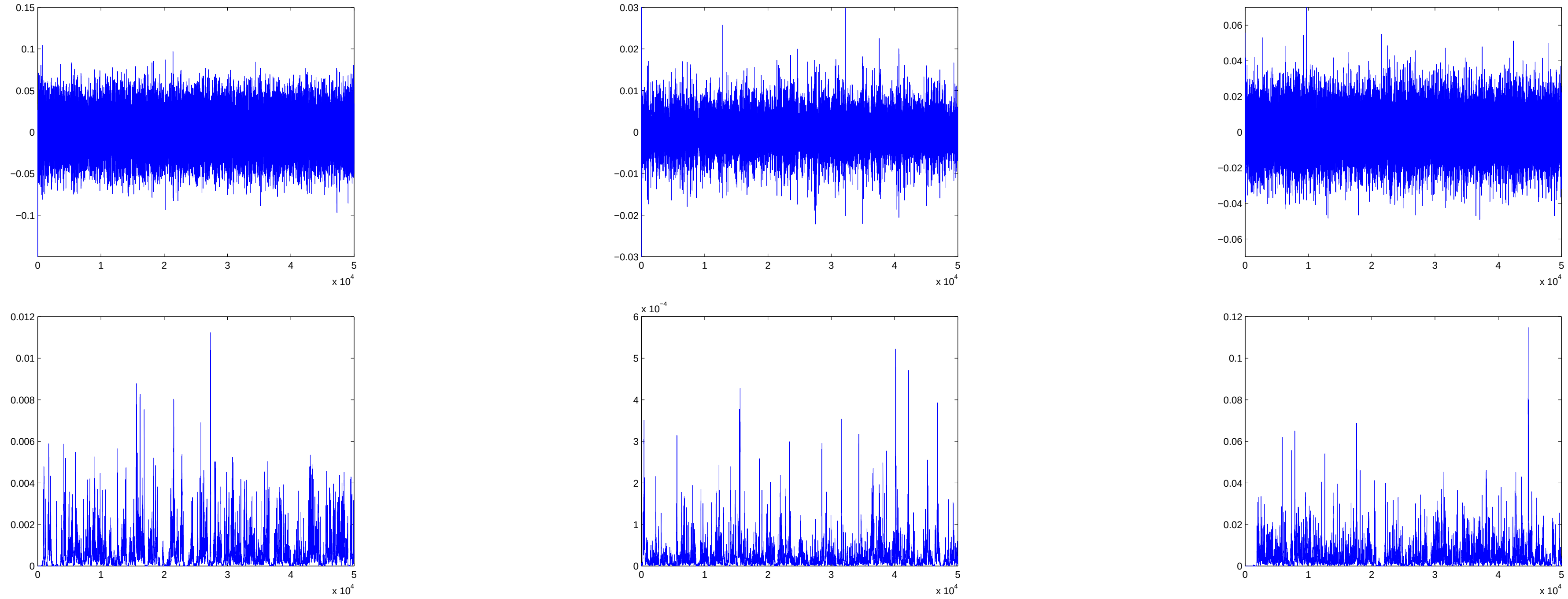


Figure 9: MCMC draws for $\pm\sqrt{\theta_1}$ (left), $\pm\sqrt{\theta_2}$ (middle) and $\pm\sqrt{\theta_3}$ (right) for a $\mathcal{N}(0,1)$-prior for $\pm\sqrt{\theta_i}, i = 1, 2, 3$ (top) and draws of $\theta_i, i = 1, 2, 3$ under a $\mathcal{G}^{-1}\left(-0.5, 10^{-7}\right)$-prior for $\theta_i, i = 1, 2, 3$ (bottom)

Variable Selection

Table 1: The two most frequently visited models (40000 MCMC iterations)

prior	δ	δ_3	γ_1	γ_2	γ_3	frequency
$p(\mu_0) \propto 1, B_0 = 1$	**0**	**1**	**0**	**1**	**0**	20331
	1	1	1	0	0	7032
$p(\mu_0) \propto 1, B_0 = 100$	**0**	**1**	**0**	**1**	**0**	26454
	0	1	1	0	0	11870
$b = 10^{-3}$	**0**	**1**	**0**	**1**	**0**	25647
	0	1	1	1	0	4173
$b = 10^{-4}$	**0**	**1**	**0**	**1**	**0**	34364
	1	1	0	1	0	1799
$b = 10^{-5}$	**0**	**1**	**0**	**1**	**0**	37012
	1	1	0	1	0	1150

Variable Selection

Table 2: Coal data; marginal posterior probability of selecting each indicator under various priors

Prior	δ	δ_3	γ_1	γ_2	γ_3
$p(\mu_0) \propto 1, B_0 = 1$	0.2375	**1.0000**	0.4131	**0.6192**	0.0597
$p(\mu_0) \propto 1, B_0 = 100$	0.0315	**1.0000**	0.3246	**0.6728**	0.0051
$b = 10^{-3}$	0.1845	**1.0000**	0.2048	**0.9295**	0.0698
$b = 10^{-4}$	0.0647	**1.0000**	0.0765	**0.9694**	0.0214
$b = 10^{-5}$	0.0347	**1.0000**	0.0386	**0.9792**	0.0078

Fixed seasonal pattern ($\delta_3 = 1, \gamma_3 = 0$), random drift ($\gamma_2 = 0$) with $a_0 = 0$ ($\delta = 0$) and $\theta_1 = 0(\gamma_1 = 0)$

Model Selection for Non-Gaussian State Space Models

Variable selection approach developed for Gaussian state space model may be extended to nonnormal state space models using auxiliary mixture sampling (Frühwirth-Schnatter and Wagner, 2006; Frühwirth-Schnatter and Frühwirth, 2007):

- state space modelling of binary time based on the logit transform
- state space modelling of categorical time based on the logit transform
- state space modelling of times of small counts based on the Poisson distribution

Illustrative application to two time series

Auxiliary Mixture Sampling for Count Data

Frühwirth-Schnatter and Wagner (2006):

- For each $y_t \sim \mathcal{P}(\lambda_t)$ introduce the hidden inter-arrival times τ_{tj}, $j = 1, \ldots, (y_t + 1)$ in the interval [0,1] of a Poisson process with intensity λ_t as missing data

- The inter-arrival times τ_{tj} are $\mathcal{E}(\lambda_t) = \mathcal{E}(1)/\lambda_t$, therefore:

$$-\log \tau_{tj} = \log \lambda_t + \varepsilon_{tj}, \quad j = 1, \ldots, (y_t + 1)$$

  where $\varepsilon_{tj} = -\log \mathcal{E}(1)$

- The distribution of ε_{tj} is approximated by a mixture of normal distributions with component indicator r_{tj} and the auxiliary variables $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_T)$, where $\mathbf{z}_t = (\tau_{tj}, r_{tj}, j = 1, \ldots, y_t + 1)$, are introduced.

Auxiliary mixture sampling and variable selection

Introducing the auxiliary variables $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_T)$, where $\mathbf{z}_t = (\tau_{tj}, r_{tj}, j = 1, \dots, y_t + 1)$, leads to a conditionally Gaussian state space model

$$\begin{aligned} -\log \tau_{tj} - \log e_t &= \mu_t + \delta_2 s_{t,0} + \gamma_3 (s_t - \delta_2 s_{t,0}) + m_{r_{tj}} + \varepsilon_t, \\ \varepsilon_t &\sim \mathcal{N}\left(0, s^2_{r_{tj}}\right). \end{aligned} \tag{9}$$

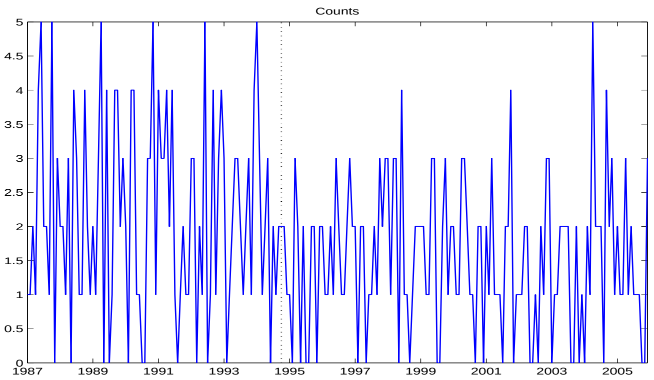
The remaining equations are the same as before. Markov chain Monte Carlo estimation is easily extended:

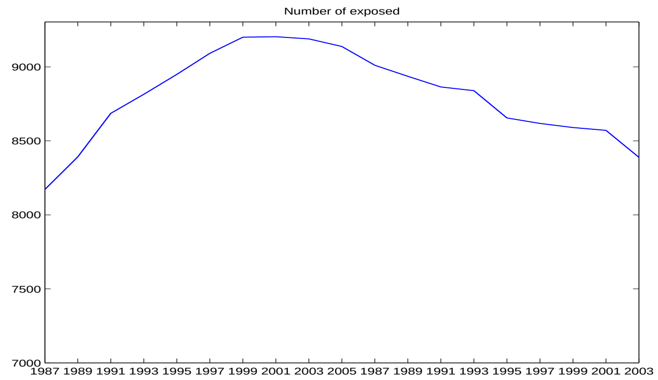
1. Variable selection and estimation for the conditionally Gaussian state space model (9) (conditional on the auxiliary variables $\mathbf{z}$)

2. Sample the auxiliary variables $\mathbf{z}$

Time Series of Road Accidents of Children

killed or injured pedestrians, Children (aged 6-10) in Linz (Austria) (left) and number of children in this age (right)
monthly data January 1987 - December 2005





A legal intervention intended to increase road safety took place during the observation period. An amendment increasing priority for pedestrians became effective on **October 1,1994**: since then pedestrians who want to use a crosswalk have to be granted crossing.

MCMC for the non-centered model ($\delta = 1$ and $\gamma = 1$)

MCMC in the centered version did not work

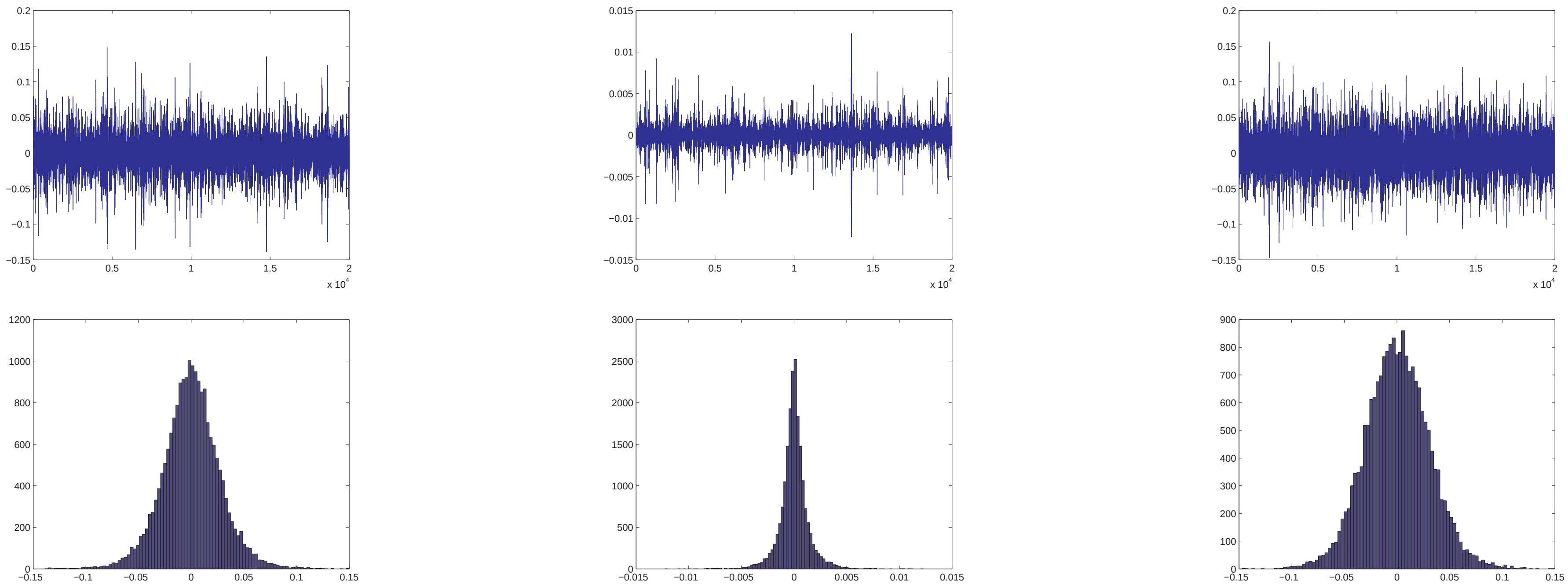


Figure 10: MCMC draws (top) and histograms (bottom) for $\pm\sqrt{\theta_1}$ (left), $\pm\sqrt{\theta_2}$ (middle) and $\pm\sqrt{\theta_3}$ (right) under the $\mathcal{N}(0, 1)$ prior for $\pm\sqrt{\theta_i}$

Variable Selection

Table 3: Two most frequently visited models (40000 MCMC iterations)

prior	δ	δ_3	δ_4	γ_1	γ_2	γ_3	frequency
$p(\mu_0) \propto 1, B_0 = 1$	**0**	**1**	**1**	**0**	**0**	**0**	37544
	0	1	1	0	0	1	937
$p(\mu_0) \propto 1, B_0 = 100$	**0**	**1**	**1**	**0**	**0**	**0**	34874
	0	1	0	0	0	0	4743
$b = 10^{-2}$	**0**	**1**	**1**	**0**	**0**	**0**	9595
	0	1	1	0	0	1	4154
$b = 10^{-3}$	**0**	**1**	**1**	**0**	**0**	**0**	18528
	1	1	0	0	0	0	4048
$b = 10^{-4}$	**0**	**1**	**1**	**0**	**0**	**0**	24871
	1	1	0	0	0	0	4717

Variable Selection

Table 4: Marginal posterior probability of selecting each indicator

	trend	season	intervention	process variances		
prior	δ	δ_3	δ_4	γ_1	γ_2	γ_3
$p(\mu_0) \propto 1, B_0 = 1$	0.0047	**1.0000**	**0.9798**	0.0209	0.0005	0.0244
$p(\mu_0) \propto 1, B_0 = 100$	0.0019	**1.0000**	**0.8767**	0.0042	0.0001	0.0035
$b = 10^{-2}$	0.3140	**1.0000**	**0.7769**	0.2872	0.2767	0.3015
$b = 10^{-3}$	0.2152	**1.0000**	**0.7094**	0.1567	0.1576	0.1289
$b = 10^{-4}$	0.1563	**1.0000**	**0.7196**	0.0772	0.0963	0.0501

Variable Selection

Simple Poisson regression model $(\gamma_1 = \gamma_2 = \gamma_3 = 0)$

- with fixed seasonal pattern $(\delta_3 = 1, \gamma_3 = 0)$
- no trend $(\delta = \gamma_2 = 0)$; fixed level before and after intervention $(\gamma_1 = 0)$
- intervention effect significant $(\delta_4 = 1)$

Variable Selection

Gain of statistical efficiency for the parameter of interest

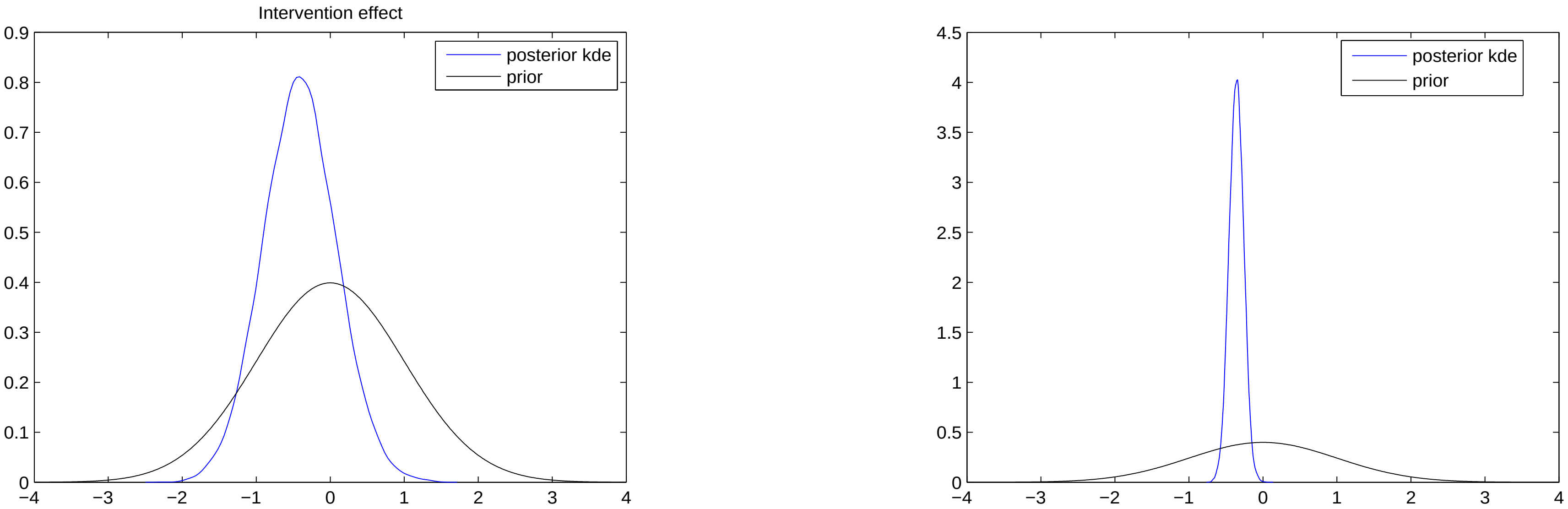


Figure 11: posterior density of the intervention effect Δ in comparison to the prior; left: unrestricted basic structural model, right: Poisson regression model with seasonal pattern (selected model)

Summary

Variable selection approach has been extended to state space models

Backbone of efficient MCMC is a non-centered parameterization in combination with a normal prior on the signed square root of the process variances

For non-Gaussian state space models the variable selection approach is feasible using auxiliary mixture sampling

Extension to other state space models (time-varying regression models, dynamic factor models) is work in progress

Summary

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

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

Chib, S. (1995). Marginal likelihood from the Gibbs output. Journal of the American Statistical Association, 90:1313–1321.

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

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. Biometrika, 89:603–615.

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

Frühwirth-Schnatter, S. (1995). Bayesian model discrimination and Bayes factors for linear Gaussian state space models. Journal of the Royal Statistical Society, Ser. B, 57:237–246.

Summary

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. The Econometrics Journal, 7:143–167.

Frühwirth-Schnatter, S. and Frühwirth, R. (2007). Auxiliary mixture sampling with applications to logistic models. Computational Statistics and Data Analysis, 51:3509–3528.

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

Frühwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of small counts with applications to state space modelling. Biometrika, 93:827–841.

Frühwirth-Schnatter, S. and Wagner, H. (2008a). Marginal likelihoods for non-Gaussian models using auxiliary mixture sampling. Computational Statistics and Data Analysis, 52:4608–4624. CHECK.

Frühwirth-Schnatter, S. and Wagner, H. (2008b). Stochastic model specification search for Gaussian and non-Gaussian state space models. Research Report IFAS 2008-36, http://www.ifas.jku.at/.

George, E. I. and McCulloch, R. (1993). Variable selection via Gibbs sampling. Journal of the American Statistical Association, 88:881–889. CHECK.

George, E. I. and McCulloch, R. (1997). Approaches for Bayesian variable selection. Statistica Sinica, 7:339–373.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika, 82:711–732.

Harvey, A. C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.

Harvey, A. C. and Durbin, J. (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. Journal of the Royal Statistical Society, Ser. A, 149:187–227. CHECK.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison (Disc: p118-138). Journal of the Royal Statistical Society, Ser. B, 57:99–118.

Tüchler, R. (2008). Bayesian variable selection for logistic models using auxiliary mixture sampling. Journal of Computational and Graphical Statistics, 17:76–94.

Zellner, A. and Rossi, P. E. (1984). Bayesian analysis of dichotomous quantal response models. Journal of Econometrics, 25:365–393. CHECK.