SFB 386 - Workshop on Model Choice and Validation
München, October 6 – 8, 2005

Model Choice and Variable Selection for Discrete Valued Data Using an Auxiliary Mixture Sampler

Sylvia Frühwirth-Schnatter,
Johannes Kepler University Linz (Austria)

joint work with
Rudi Frühwirth, Regina Tüchler, Helga Wagner, Helmut Waldl

Outline

- Auxiliary mixture sampler for Bayesian modelling of discrete-valued data
- Model choice for state space models for discrete-valued data
- Variable selection for regression models for discrete-valued data
- Covariance selection in random-effect models for discrete-valued data

Binary logit model

$$\Pr(y_i = 1|\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta})},$$

$y_i = 1/0$ . . . buy/not buy a product

$\mathbf{x}_i$ . . . product attributes (brand, price)

$\boldsymbol{\beta}$ . . . part worths

Bayesian Analysis

The presence of the logit transformation causes non-normality as well as non-linearity $\Rightarrow$ Metropolis-Hastings algorithm

A Gibbs sampling algorithm could be applied, if the model were normal and linear, in particular for a standard regression model:

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

Recalling the Concept of Utility (McFadden 1974)

Let y^u_{0i} be the utility of choosing category 0 (not buy).

Let y^u_i be the utility of consumer i to choose category 1 (buy). Utility of buying is influenced by product attributes:

$$y^u_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i.$$

If $y^u_i > y^u_{0i} \Rightarrow$ consumer will buy $\Rightarrow$ observe $y_i = 1$

If $y^u_i < y^u_{0i} \Rightarrow$ consumer will not buy $\Rightarrow$ observe $y_i = 0$

Recalling the Concept of Utility (McFadden 1974)

If y_{0i}^u and ε_i follow a type I extreme value distribution, then the marginal distribution of y_i corresponds to the binary logit regression model:

$$\Pr(y_i = 1|\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta})},$$

Data Augmentation - Step I (Scott 2004)

Estimate the latent utilities of category 1 jointly with $\boldsymbol{\beta}$ using a two step MCMC algorithm

1. Sample $\boldsymbol{\beta}$ conditional on $\mathbf{y}^u = (y_1^u, \ldots, y_N^u)$ from the model

$$y_i^u = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i. \tag{1}$$

2. Recover the latent utilities $\mathbf{y}^u = (y_1^u, \ldots, y_N^u)$ knowing $\boldsymbol{\beta}$

Model (1) is a linear regression model with non-normal error term $\varepsilon_i \Rightarrow$ Scott (2004) uses Metropolis-Hastings to sample $\boldsymbol{\beta}$

Recovering the Latent Utilities

Very efficient sampling step:

Sample the latent utility y_i^u conditional on $\boldsymbol{\beta}$ and $\mathbf{y}$ as

$$y_i^u = -\log\left(-\frac{\log(U_i)}{1+\lambda_i} - \frac{\log(V_i)}{\lambda_i} I_{\{y_i=0\}}\right),$$

where U_i and V_i are two independent uniform random numbers and $\lambda_i = \exp(\mathbf{x}_i\boldsymbol{\beta})$.

A single line of code in MATLAB or in R!

Proof requires just a bit of probability theory

Achieving Normality

To obtain a model that is conditionally normal, approximate the non-normal density $p(\varepsilon_i)$ by a normal mixture of 10 components with parameters m_r and s_r for the r-th component:

$$p(\varepsilon_i) = \exp\{-\varepsilon_i - e^{-\varepsilon_i}\} \approx \sum_{r=1}^{10} w_r f_{\mathcal{N}}(\varepsilon_i; m_r, s_r^2).$$

w_r	0.00551	0.0458	0.173	0.144	0.122	0.0989	0.102	0.114	0.106	0.0895
m_r	5.08	3.18	1.76	1.19	0.731	0.368	0.0279	- 0.314	- 0.677	- 1.06
s_r^2	3.58	1.73	0.975	0.379	0.178	0.096	0.0703	0.0705	0.0896	0.145

Achieving Normality

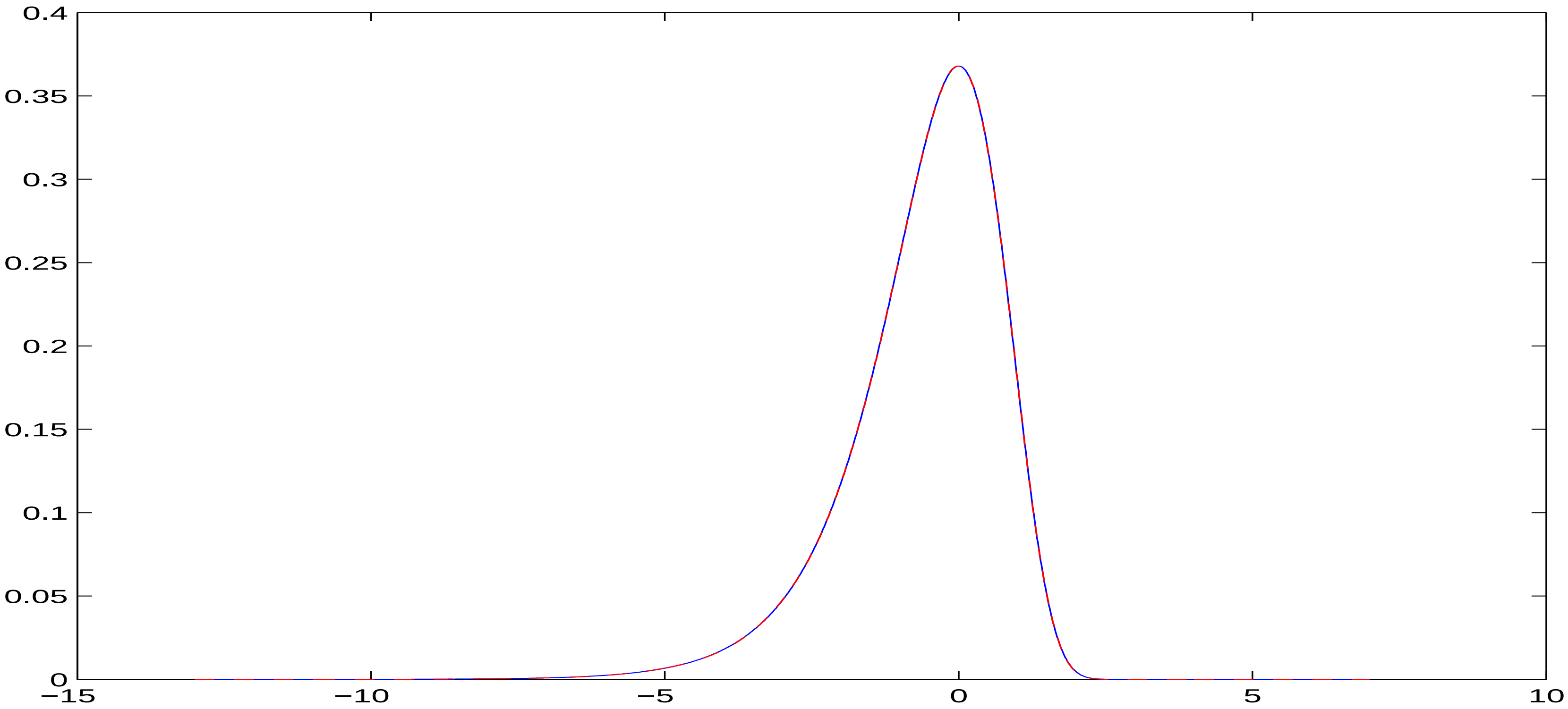

Figure 1: Comparing the density of the $-\log(\mathcal{E}(1))$-distribution (type I extreme value distribution) with a normal mixture approximation with 10 components

Data Augmentation - Step II

Frühwirth-Schnatter and Waldl (2004) - major revision with a modified list of authors coming out soon

- Introduce for each ε_i the latent component indicator r_i as missing data.

- Estimate the latent utilities y_i^u and the latent component indicator r_i jointly $\boldsymbol{\beta}$ using a two step MCMC algorithm

Auxiliary mixture sampling

Auxiliary Mixture Sampling

Estimate the latent utilities y_i^u and the latent component indicator r_i jointly $\boldsymbol{\beta}$ using a two step MCMC algorithm

1. Sample $\boldsymbol{\beta}$ conditional on $(y_1^u, \ldots, y_N^u)$ and indicator $(r_1, \ldots, r_N)$ from the normal, linear regression model

$$y_i^u = \mathbf{x}_i \boldsymbol{\beta} + m_{r_i} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}\left(0, s_{r_i}^2\right).$$

2. NEW: Recover the latent utilities y_i^u and the latent indicators r_i conditional on knowing $\boldsymbol{\beta}$

Auxiliary Mixture Sampling - Step 2

- Sample the latent utility y_i^u conditional on $\boldsymbol{\beta}$ and $\mathbf{y}$ as explained above (need two uniform random number and $\lambda_i = \exp(\mathbf{x}_i\boldsymbol{\beta})$)

- Sample r_i conditional on knowing y_i^u from the following discrete density, where $(w_j, m_j, s_j^2), j = 1, \ldots, 10$ are the parameters of the finite mixture approximation:

$$\Pr(r_i = j | y_i^u, \boldsymbol{\beta}) \propto \frac{w_j}{s_j} \exp(-\frac{(y_i^u - \log \lambda_i - m_j)^2}{2s_j^2}).$$

Related Work

Related algorithm with different kind of data augmentation:
Holmes and Held (2006)

Related work using mixtures approximation:

- Stochastic volatility modelling: Kim, Shephard, and Chib (1998) and Chib, Nardari, and Shephard (2002) use a 7 component normal mixture approximation of the density of a $\log \chi^2$-distribution

- Spectral analysis: Carter and Kohn (1997) use a 5 component normal mixture approximation of the density of a $\log(\mathcal{E}(1))$-distribution

Multinomial data

- Data with $m+1$ categories, labelled $\{0,1,\ldots,m\}$
- There is a latent utility $y^u_{ki}, k=0,\ldots,m$ for each category. $y^u_{ki}, k \geq 1$ is modelled as being dependent on covariates:

$$y^u_{ki} = \mathbf{x}_i \boldsymbol{\beta}_k + \varepsilon_{ki}.$$

- The observed category is the category with maximal utility
- The multinomial model is the marginal distribution of y_i, if y^u_{0i} and $\varepsilon_{ki}, k=1,\ldots,m$, follow a type I extreme value distribution (McFadden 1974).

Auxiliary Mixture Sampling for Multinomial Data

1. Draw the parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$ from the normal posterior density of following standard regression model:

$$y_{ki}^u = \mathbf{x}_i \boldsymbol{\beta}_k + m_{r_{ki}} + \varepsilon_{ki}, \quad \varepsilon_{ki} \sim \mathcal{N}\left(0, s_{r_{ki}}^2\right).$$

2. NEW: Recover the utility y_{ki}^u and the indicator r_{ki} for each $i = 1, \ldots, N$ and for each $k = 1, \ldots, m$ using $\log \lambda_{ki} = \mathbf{x}_i \boldsymbol{\beta}_k$

Holmes and Held (2006) - sample $\boldsymbol{\beta}_k | \boldsymbol{\beta}_{-k}$

Regression Models for Count Data

$$y_i|\boldsymbol{\beta} \sim \mathcal{P}\left(\exp(\mathbf{x}_i\boldsymbol{\beta})\right) = \mathcal{P}\left(\lambda_i\right)$$

y_i . . . observed counts

$\mathbf{x}_i$ . . . covariates

$\boldsymbol{\beta}$ . . . regression coefficient

Auxiliary Mixture Sampling for Count Data

Frühwirth-Schnatter and Wagner (2004)

- Instead of the utilities use the inter-arrival times of a hidden Poisson process with mean λ_i

- Mixture approximation of the density of a type I extreme value distributed random variable: Introduction of the component indicator of this mixture as a second sequence of missing data

Excursion: The Poisson Process

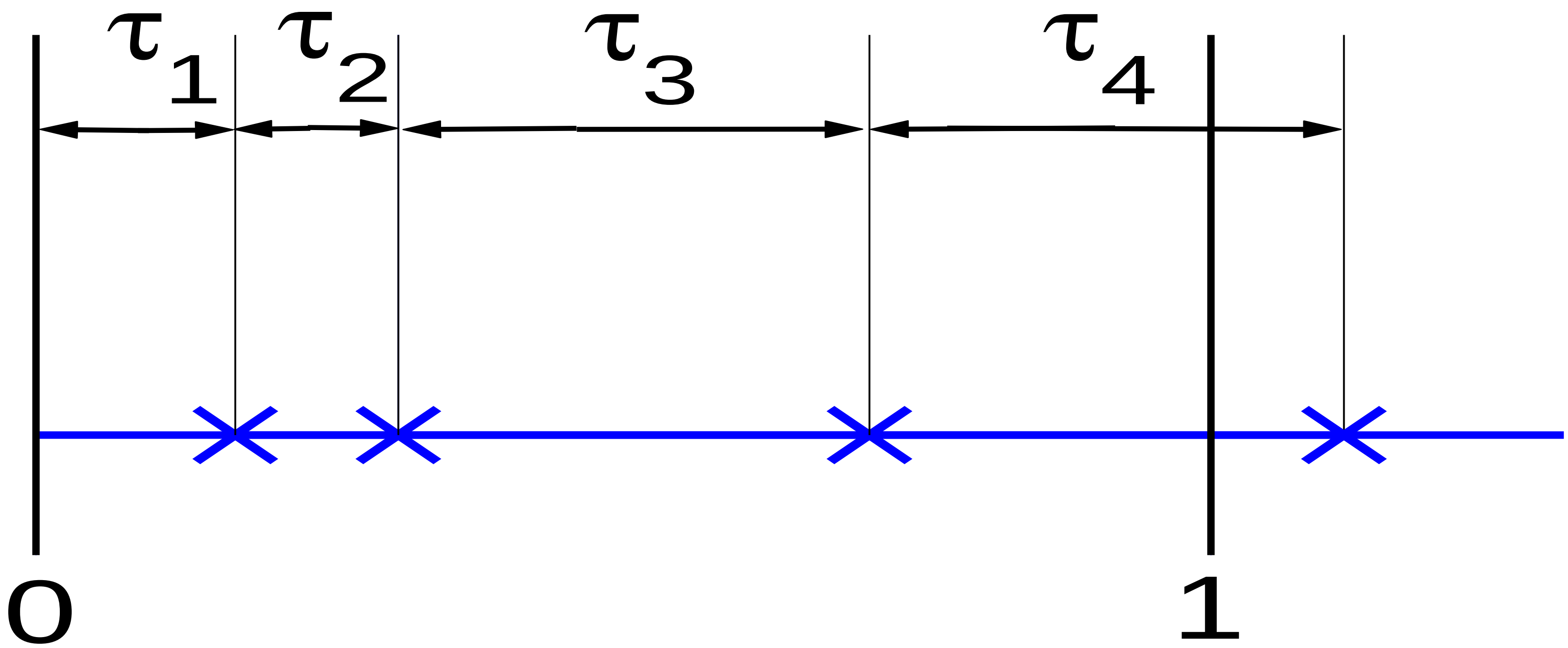

Figure 2: Poisson process: crosses mark the occurrence of an event

Achieving Linearity - Data Augmentation Step 1

- For each $i = 1, \ldots, N$ introduce the hidden inter-arrival times τ_{ij}, $j = 1, \ldots, (y_i + 1)$, of the unobserved Poisson process as missing data

- The inter-arrival times τ_{ij} are $\mathcal{E}(\lambda_t)$ distributed, therefore:

$$-\log \tau_{ij} = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_{ij}, \quad j = 1, \ldots, (y_i + 1)$$

$p(\varepsilon_{ij})$ type I extreme value distribution

Achieving Normality - Data Augmentation Step 2

- Same as for binary and multinomial data

- Approximation of the density of ε_{ij} by a normal mixture of 10 components $\Rightarrow$ indicator r_{ij} for each $i = 1, \ldots, N$ and for each $j = 1, \ldots, (y_i + 1)$

Auxiliary Mixture Sampling for Count Data

1. Sample $\boldsymbol{\beta}$ conditional on all τ_{ij} and r_{ij} from the normal posterior of following standard regression model:

$$-\log \tau_{ij} = \mathbf{x}_i\boldsymbol{\beta} + m_{r_{ij}} + \varepsilon_{ij}, \quad \varepsilon_i \sim \mathcal{N}\left(0, s^2_{r_{ij}}\right).$$

2. NEW: Recover the inter arrival times τ_{ij} and the indicator r_{ij} for each $i = 1, \ldots, N$ and for each $j = 1, \ldots, (y_i + 1)$ using $\log \lambda_{ij} = \mathbf{x}_i\boldsymbol{\beta}$

More Complex Models

Auxiliary mixture sampling only based on knowing $\lambda_i = \exp(\mathbf{x}_i\boldsymbol{\beta}) \Rightarrow$ easily extended to more general models:

- State-space models Frühwirth-Schnatter and Wagner (2004)
- Random-effect models (Frühwirth-Schnatter and Waldl (2004), Tüchler (2003))
- Markov switching autoregressive models for binary time series and times series of counts (Stefan, 2005)
- Space-time models (Gschlössl and Czado,2005)

Application to Model Choice and Variable Selection

- Model choice for state space models for discrete-valued data
- Variable selection for regression models for discrete-valued data
- Covariance selection in random-effect models for discrete-valued data

Time Series of Road Accidents of Pedestrians

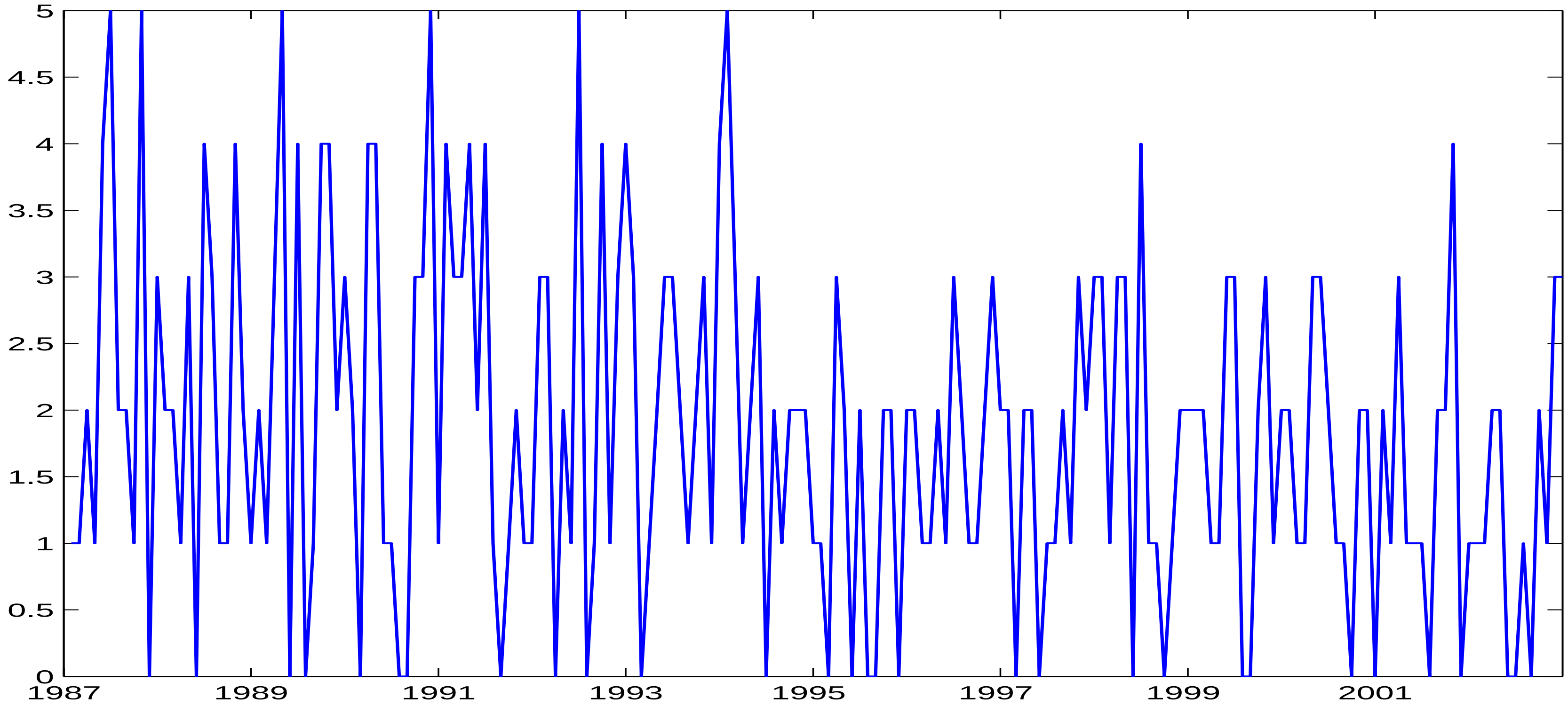

Figure 3: Killed or seriously children (aged 6-10) in Linz (Austria), 1987:01 - 2002:12

Time Series of Road Accidents of Pedestrians

A legal intervention intended to increase road safety took place during the observation period. More precisely, 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.

$\Rightarrow$ Study intervention effect

Basic structural model for Poisson counts

The intensity λ_t has a trend and a seasonal component.

$$y_t \sim \mathcal{P}\left(e^{\mu_t} e^{s_t}\right)$$

$$y_t \sim \mathcal{P}\left(\lambda_t\right), \quad \log(\lambda_t) = \mu_t + s_t$$

Basic structural model for Poisson counts

- Trend component

$$\mu_t = \mu_{t-1} + a_{t-1} + w_{1t}, \qquad w_{1t} \sim \mathcal{N}(0, \theta_1)$$

$$a_t = a_{t-1} + w_{2t}, \qquad w_{2t} \sim \mathcal{N}(0, \theta_2)$$

- Intervention effect: Modification of trend component for $t = t_{int}$:

$$\mu_t = \mu_{t-1} + a_{t-1} + \delta + w_{1t}$$

- seasonal component

$$s_t = -s_{t-1} - \cdots - s_{t-11} + w_{3t}, \qquad w_{3t} \sim \mathcal{N}\left(0, \theta_4^2\right)$$

Dynamic Generalized Linear Model

Special case of following state space form for Poisson counts:

$$y_t|\lambda_t \sim \mathcal{P}(\lambda_t)$$

$$\lambda_t = \exp(Z_t^1 \alpha + Z_t^2 \beta_t)$$

$$\beta_t = F_t \beta_{t-1} + w_t \qquad w_t \sim \mathcal{N}(0, Q)$$

Joint prior distribution $p(\beta^T)$ of the state vector $\beta^T = (\beta_0, \ldots, \beta_T)$ and $p(\alpha)$ are normal; $p(Q)$ is an inverted Wishart or the product of inverted Gammas

Auxiliary Mixture Sampling for DGLM

- Multi-move sampling of the whole sequence $\beta_0, \ldots, \beta_T$ and α by forward filtering-backward sampling from $p(\alpha, \beta^T | y, Q, S, \tau)$ as in (Frühwirth-Schnatter 1994), (Carter and Kohn 1994) and (De Jong and Shephard 1995): recursive sampling from multivariate normals with rank(Q)

- sample Q from $p(Q|y, \beta^T, R, \tau)$ (inverted Wishart distribution).

- NEW: Joint sampling of R, τ from $p(R, \tau | y, \beta^T, Q)$

Auxiliary Mixture Sampling for Road Safety Data

- Auxiliary mixture sampling was run 12000 times with a burn-in of 2000 runs.

- Priors: $\mu_0 \sim \mathcal{N}(0,1)$, $a_0 \sim \mathcal{N}(0,1)$, $s_j \sim \mathcal{N}(0,1)$, $j = 1, \ldots, 12$.

Time Series of Road Accidents of Children

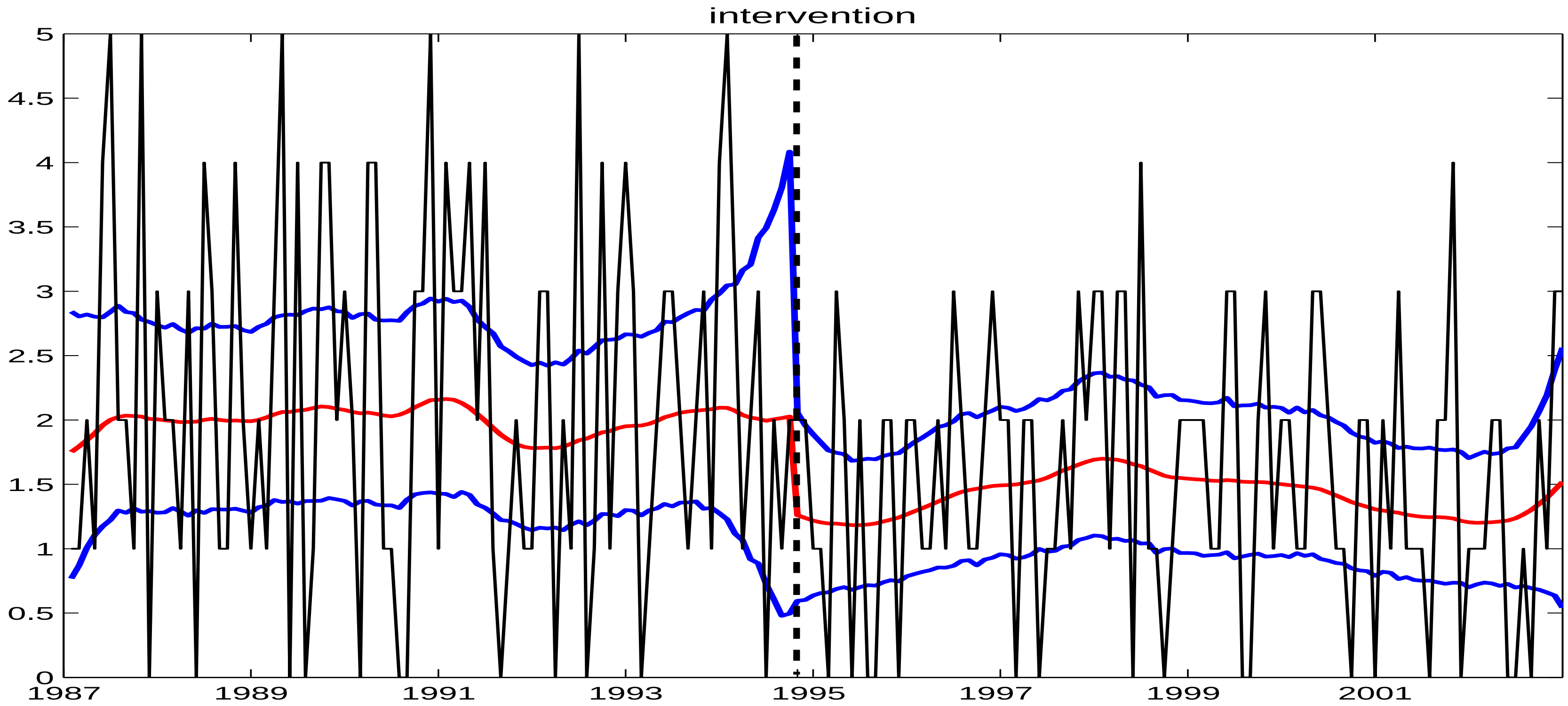

Figure 4: Counts and estimated trend component $\exp(\mu_t)$

Time Series of Road Accidents of Children

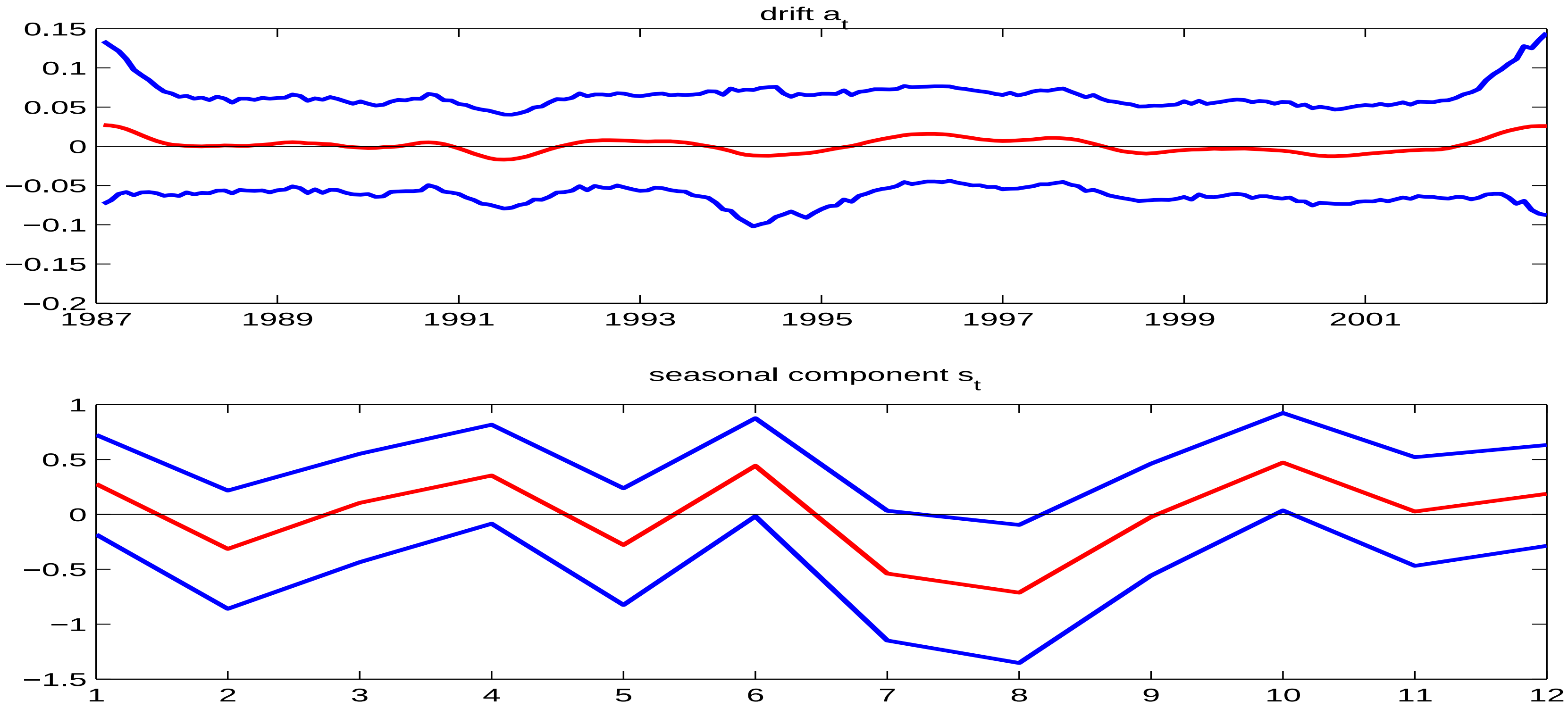

Figure 5: Trend and seasonal component

Time Series of Road Accidents of Children

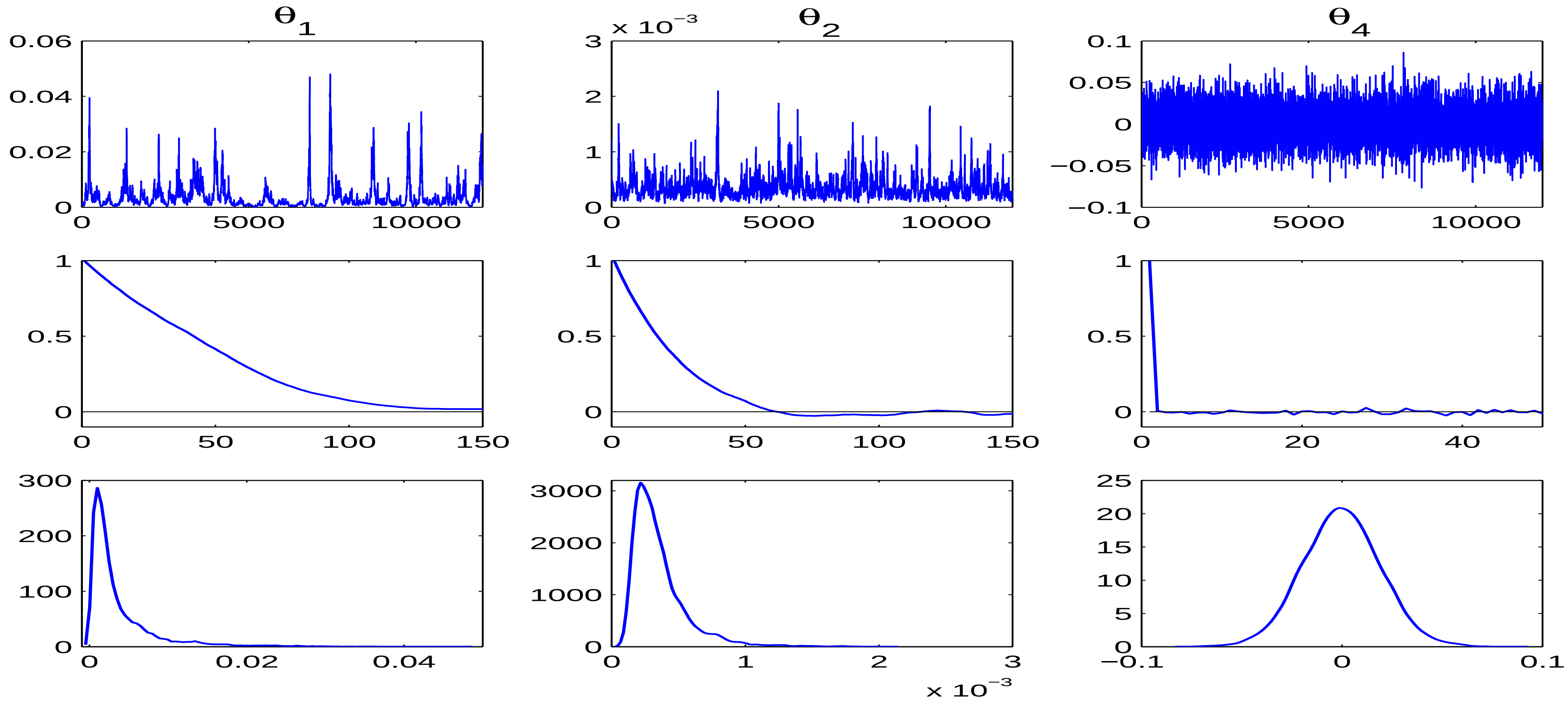

Figure 6: Auxiliary Gibbs sampling results for the variances

Model Choice for Using Marginal Likelihoods

Estimation of marginal likelihoods for state space models of discrete-valued data by the method of Chib (1995) is feasible from the output of the auxiliary mixture sampler:

$$p(y) = \frac{p(y|(\beta_0, \beta_1, \dots, \beta_T, \alpha)^*)p((\beta_0, \beta_1, \dots, \beta_T, \alpha)^*)}{p((\beta_0, \beta_1, \dots, \beta_T, \alpha)^*|y)}$$

Model Choice for Using Marginal Likelihoods

$p((\beta_0, \beta_1, \ldots, \beta_T, \alpha)^*|y)$ is computed from the output of the auxiliary mixture sampler as following mixture of normal distributions

$$p((\beta_0, \beta_1, \ldots, \beta_T, \alpha)^*|y)$$
$$= \frac{1}{M} \sum_{m=1}^{M} p((\beta_0, \beta_1, \ldots, \beta_T, \alpha)^*|\boldsymbol{\tau}^{(m)}, R^{(m)}, \boldsymbol{\vartheta}^{(m)}, y)$$

Similar formula for binary and multinomial data

Application to Road Safety Data

Table 1: Marginal Likelihoods

local level model - fixed season, no intervention	-242.6029
local level model - fixed season, intervention	**-191.5628**
dynamic trends model - fixed season, no intervention	-271.8189
dynamic trends model - fixed season,intervention	-258.9480
basic structural model, intervention	–
basic structural model, no intervention	–
fixed effects - full seasonal, intervention	-329.3423
fixed effects - full seasonal, no intervention	-331.5685
fixed effects - holiday effect, intervention	-308.9187

Variable selection for discrete-valued regression models

y_i . . . dependent for subject i, $i = 1, \ldots, N$
Z_i . . . design matrix / covariates $(1 \times d)$
$\boldsymbol{\beta}$ . . . regression parameter

Regression model for discrete-valued data

$$\Pr(y_i = 1|\boldsymbol{\beta}) = \frac{\exp(Z_i\boldsymbol{\beta})}{1 + \exp(Z_i\boldsymbol{\beta})},$$
$$y_i \sim \mathcal{P}\left(\exp(Z_i\boldsymbol{\beta})\right)$$

Extension to multinomial regression models possible

Variable selection for discrete-valued regression models

Variable Selection for the elements of $\boldsymbol{\beta}$ as for the standard linear regression model, see e.g. George and McCulloch (1997)

We define indicators δ_j for each element β_j of $\boldsymbol{\beta}$, $j = 1, \ldots, d$:

$$\beta_j = 0, \quad \text{iff} \;\; \delta_j = 0,$$
$$\beta_j \neq 0, \quad \text{iff} \;\; \delta_j = 1.$$

Related work by Holmes and Held (2006) for the binary and the multinomial regression model

Variable selection through auxiliary mixture sampling

Introducing the auxiliary variables $\mathbf{R}$ (indicators in the mixture) and $\boldsymbol{\tau}$ ($\mathbf{y}^u$) leads to a conditionally Gaussian model

1. Sampling of the indicators $\boldsymbol{\delta}$ recursively from $p(\delta_j|\boldsymbol{\delta}_{-j}, \mathbf{R}, \boldsymbol{\tau}, \mathbf{y})$, $j = 1, \ldots, d$;

2. Sampling of the non-zero regression parameters $\boldsymbol{\beta}^{\boldsymbol{\delta}}$ from the normal posterior distribution $p(\boldsymbol{\beta}^{\boldsymbol{\delta}}|\boldsymbol{\delta}, \mathbf{R}, \boldsymbol{\tau}, \mathbf{y})$

3. NEW: Joint sampling of $(\mathbf{R}, \boldsymbol{\tau})$ from $p(\mathbf{R}, \boldsymbol{\tau}|\mathbf{y}, \boldsymbol{\beta})$ using $\log \lambda_i = Z_i\boldsymbol{\beta}, i = 1, \ldots, N$

Example - Credit-scoring data

Fahrmeir, Hamerle, Tutz (1996)

- 1000 consumer credits:
  $y_{it} = 1$ for 700 creditworthy,
  $y_{it} = 0$ for 300 not creditworthy

- 20 risk factors are observed; e.g. duration of credit, amount of credit, intended use, marital status, age, ...

We select risk factors and estimate the effect of these risk factors.

Example - Credit-scoring data

- We use dummy coding for the categorical variables and have 36 elements in θ.

- We start with the most general model, which includes all effects.

- We carry out 50.000 iterations of the auxiliary mixture sampler and leave 10.000 for burn-in.

Example - Credit-scoring data

Posterior probability of the indicators to be non-zero ($M = 40000$):

$$\hat{P}(\delta_j = 1|y) = \frac{1}{S} \sum_{m=1}^{M} \delta_j^{(m)},$$

12 effects with posterior probability greater than 0.5.

Example - Credit-scoring data

risk-factor	$\hat{P}(\delta = 1 \vert y) = 1$	$\hat{\theta}$
no running account	0.9	-0.54
good running account	1	1.3
duration	0.99	-0.04
credit-worthy in the past	0.99	0.96
private purpose	0.96	0.53
higher savings	0.98	0.81
same employer		
≤ 1 year	0.44	-0.24
$1 < .. \leq 4$ years	0.22	-0.02
> 4 years	0.67	0.33

Example - Credit-scoring data

risk-factor	$\hat{P}(\delta = 1 \mid y) = 1$	$\hat{\theta}$
rate (% of income)		
$20 \leq .. < 35$	0.17	-0.00
< 20	0.83	-0.44
male, not single	0.52	0.21
female, single	0.12	0.02
other debtors	0.19	-0.09
surety	0.67	0.66
same home		
$1 < .. \leq 7$ years	0.43	-0.18
$.. > 7$ years	0.16	-0.02

Example - Credit-scoring data

risk-factor	$\hat{P}(\delta = 1 \vert y) = 1$	$\hat{\theta}$
no foreign worker	0.82	1.28
car owner	0.14	-0.02
life insurance	0.11	-0.01
real estate	0.26	-0.12
age	0.37	0.00
other credits		
bank	0.11	-0.03
others	0.67	0.33
rented flat	0.23	-0.08
freehold flat	0.15	0.04

Example - Credit-scoring data

risk-factor	$\hat{P}(\delta = 1\|y) = 1$	$\hat{\theta}$
no. of credits in the past		
$1, 3$	0.11	0.01
≥ 4	0.13	0.05
unskilled, resident	0.1	-0.00
skilled	0.11	-0.01
manager/self-employed	0.1	-0.00
> 3 persons entitled to		
maintenance	0.12	0.01
telephone	0.24	0.06
intercept		-0.29

Covariance Selection for Random Effect Models

Chen and Dunson (2003), Frühwirth-Schnatter and Tüchler (2004): Covariance selection for random effect models for normally distributed data

- Decompose the covariance matrix of the random effects using the Cholesky decomposition

- Use the non-centered parameterization $\longrightarrow$ the Cholesky factors are regression coefficients in a regression model

- Use variable selection on the elements of the Cholesky factors

Binary logit random effects model

Binary data $y_{it} \in \{0, 1\}$, $i = 1, \ldots, N$ subjects, $t = 1, \ldots, T_i$ repetitions

$$Pr(y_{it} = 1|\boldsymbol{\beta}_i) = \frac{\exp(x_{it}\boldsymbol{\beta}_i)}{1 + \exp(x_{it}\boldsymbol{\beta}_i)},$$

$$\boldsymbol{\beta}_i \sim N_{d_r}(\boldsymbol{\beta}^G, \mathbf{Q}).$$

Covariance Selection

Tüchler (2003): extension to discrete-valued data is possible using the auxiliary mixture sampler

- Cholesky decomposition: $Q = C \cdot C'$ (C lower triangular)
- $\beta_i = \beta^G + C \cdot \tilde{z}_i$ with $\tilde{z}_i \sim N_{d_r}(0, I)$.

$$Pr(y_{it} = 1|\beta^G, C, \tilde{z}_i) = \frac{\exp(x_{it}\beta^G + x_{it}C\tilde{z}_i)}{1 + \exp(x_{it}\beta^G + x_{it}C\tilde{z}_i)}$$

$\longrightarrow$ Variable selection for the elements of C.

Covariance Selection

Define indicators γ_{lm} for the lower triangular elements C:

$$\begin{aligned} C_{lm} &= 0, \quad \text{iff} \;\; \gamma_{lm} = 0, \\ C_{lm} &\neq 0, \quad \text{iff} \;\; \gamma_{lm} = 1, \\ \text{for} \;\; l &\geq m \end{aligned}$$

$\longrightarrow$ Estimate the non-zero elements C^γ from the reduced model:

$$Pr(y_{it} = 1|\beta^G, C, \boldsymbol{\gamma}, \tilde{z}_i) = \frac{\exp(x_{it}\beta^G + x_{it}^\gamma C^\gamma \tilde{z}_i)}{1 + \exp(x_{it}\beta^G + x_{it}^\gamma C^\gamma \tilde{z}_i)}.$$

Covariance selection through auxiliary mixture sampling

1. Sampling of the indicators $\boldsymbol{\gamma}$ recursively from $p(\boldsymbol{\gamma}_{lm}|\boldsymbol{\gamma}_{-lm}, \mathbf{R}, \boldsymbol{\tau}, \mathbf{y}, z^N)$;

2. Sampling of the regression parameter $\boldsymbol{\beta}^G$ and the Cholesky factor C^{γ} from the normal posterior distribution $p(\boldsymbol{\beta}^G, C^{\gamma}|\boldsymbol{\gamma}, \mathbf{R}, \boldsymbol{\tau}, z^N, \mathbf{y})$

3. Sampling of the random effects $\tilde{z}^N = \{\tilde{z}_i\}$ from $p(z^N|\boldsymbol{\beta}^G, \mathbf{R}, \boldsymbol{\tau}, \mathbf{y})$;

4. NEW: Joint sampling of $(\mathbf{R}, \boldsymbol{\tau})$ from $p(\mathbf{R}, \boldsymbol{\tau}|\mathbf{y}, \boldsymbol{\beta})$ using $\log \lambda_{it} = x_{it}\beta_i$, $i = 1, \ldots, N$, $t = 1, \ldots, T_i$

Travel Data - Marketing Application

- Conjoint study about packaged city trips (Hatzinger and Mazanez, 2005)

- 499 consumers were asked to rate 9 different city trip packages (1 ... high, 0 ... low)

Attributes:

destination:	close to Vienna or medium distance
transport mode:	bus ; train; plane
duration of stay:	2 days; 3 days; 4 days
price in Euro:	258; 372 ; 482

Auxiliary Mixture Sampling and Covariance Selection

- Fractional prior for $\boldsymbol{\beta}^G$ and $C^{\boldsymbol{\gamma}}$;
- $M = 40000$ iterations (after 10000 iterations burn-in).

Travel Data - estimates of β^G

baseline: close to Vienna, bus, 2 days, 258 Euro

baseline	-1.8
medium distance to Vienna	0.8
transport mode - train	1.1
transport mode - plane	3.0
duration of stay - 3 days	0.7
duration of stay - 4 days	0.8
price in Euro 372	-0.6
price in Euro 482	-2.2

Travel Data - estimates of indicators for Q

1	0.95	1	1	0.83	0.56	0.43	0.39
0.95	1	1	0.99	0.86	0.89	0.61	0.63
1	1	1	1	0.9	0.92	0.74	0.89
1	0.99	1	1	0.91	0.9	0.77	0.88
0.83	0.86	0.9	0.91	1	1	1	1
0.56	0.89	0.92	0.9	1	1	1	1
0.43	0.61	0.74	0.77	1	1	1	1
0.39	0.63	0.89	0.88	1	1	1	1

Summary

- Computational device for model selection and variable selection for discrete-valued data
- Data augmentation using auxiliary variables lead to a conditionally Gaussian models
- No free lunch: there is a small approximation error when using the auxiliary mixture sampler

References

*References

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

Carter, C. K. and R. Kohn (1997). Semiparametric Bayesian inference for time series with mixed spectra. Journal of Royal Statistical Society, Series B 59, 255–268.

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

Chib, S. (1995). Marginal likelihood from the Gibbs output.

References

Journal of the American Statistical Association 90, 1313–1321.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics 108, 281–316.

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

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

Frühwirth-Schnatter, S. and R. Tüchler (2004). Bayesian parsimonious covariance estimation for hierarchical linear

References

mixed models.

Frühwirth-Schnatter, S. and H. Wagner (2004). Gibbs sampling for parameter-driven models of time series of small counts with applications to state space modelling. Research Report IFAS, http://www.ifas.jku.at/.

Frühwirth-Schnatter, S. and H. Waldl (2004). Data augmentation and Gibbs sampling for logistic models. Research Report IFAS 2004-06, http://www.ifas.jku.at/.

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

Holmes, C. C. and L. Held (2006). Bayesian auxiliray variable

References

models for binary and multinomial regression. Bayesian Analysis 1, 145–168.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393. CHECK.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), Frontiers of Econometrics, pp. 105–142. New York: Academic. CHECK.

Scott, S. L. (2004). Data augmentation, frequentistic estimation, and the Bayesian analysis of multinomial logit models.

Tüchler, R. (2003). Bayesian Modelling of Unobserved

References

Heterogeneity - with an Application to Metric Conjoint Analysis. Phd thesis, University of Technology, Vienna, Austria.