Model Selection for Non-Gaussian Data Based on Auxiliary Mixture Sampling

Sylvia Frühwirth-Schnatter
Johannes Kepler Universität Linz

joint work with Regina Tüchler and Helga Wagner

Outline

- The Bayesian approach to model selection
- Auxiliary mixture sampling – simple MCMC for non-Gaussian models
- Variable selection in a Poisson regression model using marginal likelihoods
- Variable selection in a Poisson regression model using stochastic search variables
- Model specification search for state modeling of count data

The Bayesian Approach

Consider K candidate models $\mathcal{M}_1, \dots, \mathcal{M}_K$, defined by the likelihood $p(\mathbf{y}|\boldsymbol{\vartheta}_k, \mathcal{M}_k)$ and the prior $p(\boldsymbol{\vartheta}_k|\mathcal{M}_k)$.

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

Compute the posterior probability distribution over the model space, i.e. $p(\mathcal{M}_k|\mathbf{y}), k = 1, \dots K$ 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 $\mathcal{M}_k$:

$$p(\mathbf{y}|\mathcal{M}_k) = \int_{\Theta_k} p(\mathbf{y}|\boldsymbol{\vartheta}_k, \mathcal{M}_k)p(\boldsymbol{\vartheta}_k|\mathcal{M}_k)d\boldsymbol{\vartheta}_k.$$

Challenges with the Bayesian Approach

- Computing marginal likelihoods – this is a numerical challenge as the integration may be high-dimensional: importance sampling (Zellner and Rossi, 1984; Frühwirth-Schnatter, 1995), Chib's estimator (Chib, 1995), bridge sampling (Meng and Wong, 1996; Frühwirth-Schnatter, 2004)

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

- Prior distributions – prior distributions have to be proper (O'Hagan, 1995), priors on parameters appearing only in some of the models may be influential (Lindley, 1957)

Auxiliary mixture sampling

Auxiliary mixture sampler leads to simple, straightforward MCMC estimation for a Bayesian analysis of rather general parameter-driven models of discrete-valued data like random effect models, mixture models, spatial models or state space models:

- count data from the Poisson distribution (Frühwirth-Schnatter and Wagner, 2006b,a; Frühwirth-Schnatter, Frühwirth, Held, and Rue, 2007)
- binary and multinomial data based on the logit transform (Holmes and Held, 2006; Frühwirth-Schnatter and Frühwirth, 2007)
- observations from the binomial distribution (Frühwirth-Schnatter and Frühwirth, 2007; Frühwirth-Schnatter, Frühwirth, Held, and Rue, 2007)

Implementing Bayesian Model Selection for non-Gaussian Data

Auxiliary mixture sampling facilitates the implementation of Bayesian model selection for parameter-driven models of discrete-valued data:

- Marginal likelihood computation for regression models for non-Gaussian data (Frühwirth-Schnatter and Wagner, 2007a)

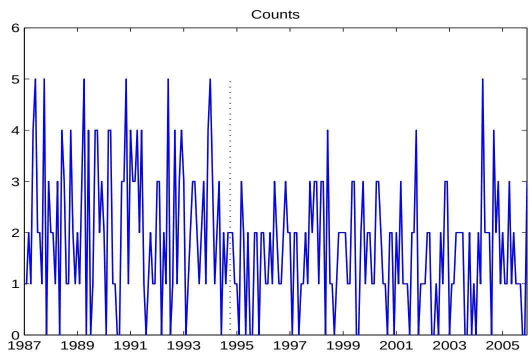
- Variable selection in regression models for binary data (Holmes and Held, 2006; Tüchler, 2007), extension to count data and other discrete-valued data possible

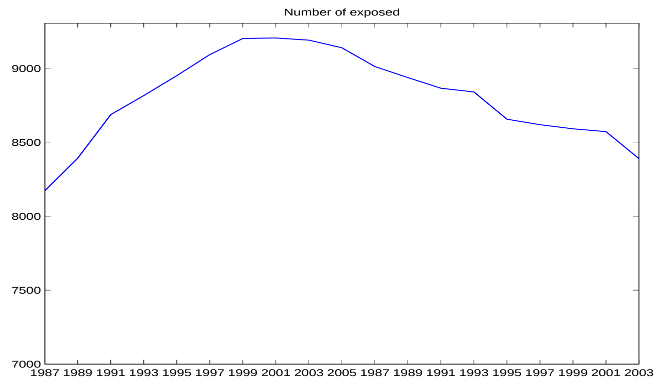
- Covariance selection in random-effects models for binary data (Tüchler, 2007), extension to count data and other discrete-valued data possible

- Stochastic specification search for Gaussian and non-Gaussian state space models (Frühwirth-Schnatter and Wagner, 2007b)

Example: 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 (increased priority for pedestrians) intended to increase road safety became effective on **October 1,1994**

Example: Time Series of Road Accidents of Children

What effects are present: seasonal effects? trend effects? intervention effect?

$\Rightarrow$ Use marginal likelihoods and stochastic search variables for variable selection in a simple Poisson regression models

Harvey and Durbin (1986) suggest to use state space modeling for road safety data – are these effects dynamic or not?

$\Rightarrow$ Extend the stochastic search variables approach to test for dynamic components in a Poisson state space model

Parameter-driven Modelling of Count Data

Count data $\mathbf{y} = (y_1, \ldots, y_T)$ (single measurements, time series, spatial data, repeated measurements, panel data, etc.)

Parameter-driven model:

$$y_t \sim \mathcal{P}\left(e_t \lambda_t\right).$$

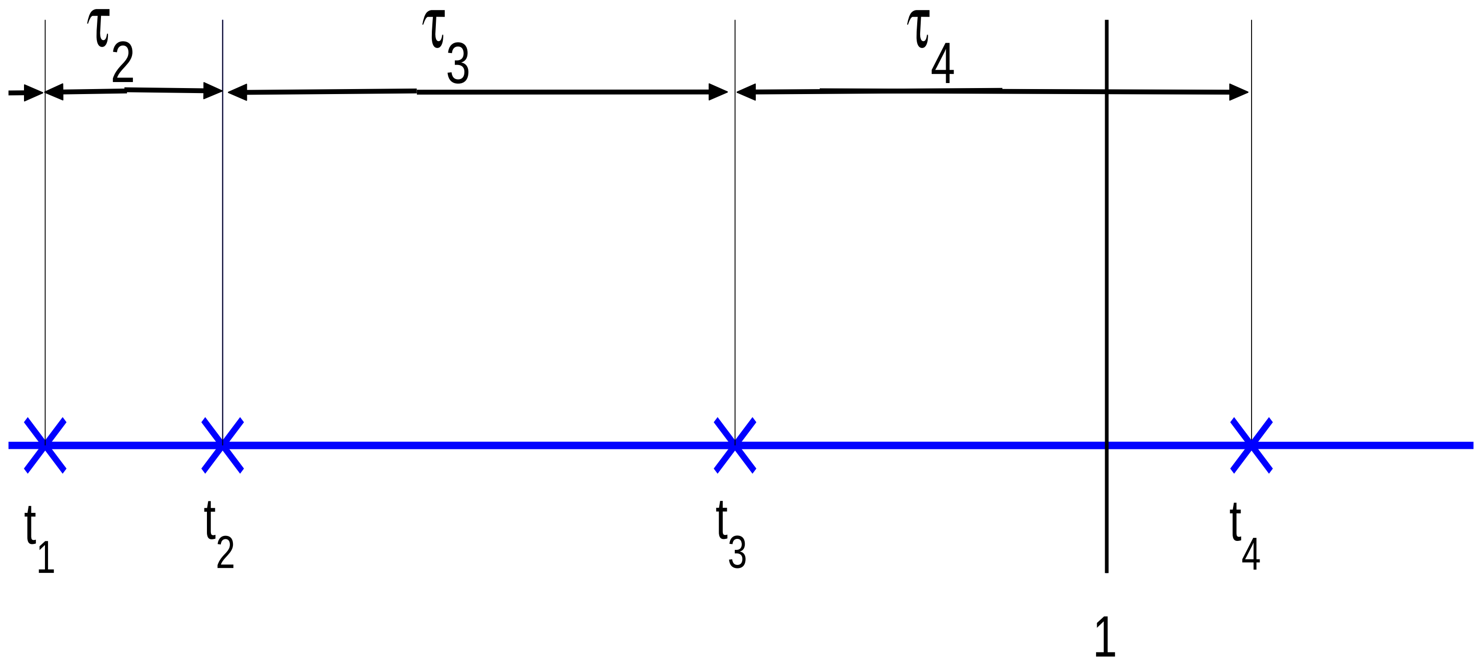
- Regression model with regression coefficient $\boldsymbol{\beta}$:

$$\log \lambda_t = \mathbf{x}_t \boldsymbol{\beta}$$

- State space model, e.g. local level model

$$\log \lambda_t = \beta_t, \qquad \beta_t = \beta_{t-1} + \omega_t, \quad \omega_t \sim \mathcal{N}\left(0, \theta\right)$$

Properties of the Poisson Process



Poisson process: crosses mark the occurrence of an event

$t_1, \ldots t_4$ are the arrival times of process,

$\tau_1, \ldots \tau_4$ are inter-arrival times

Mixture Approximation

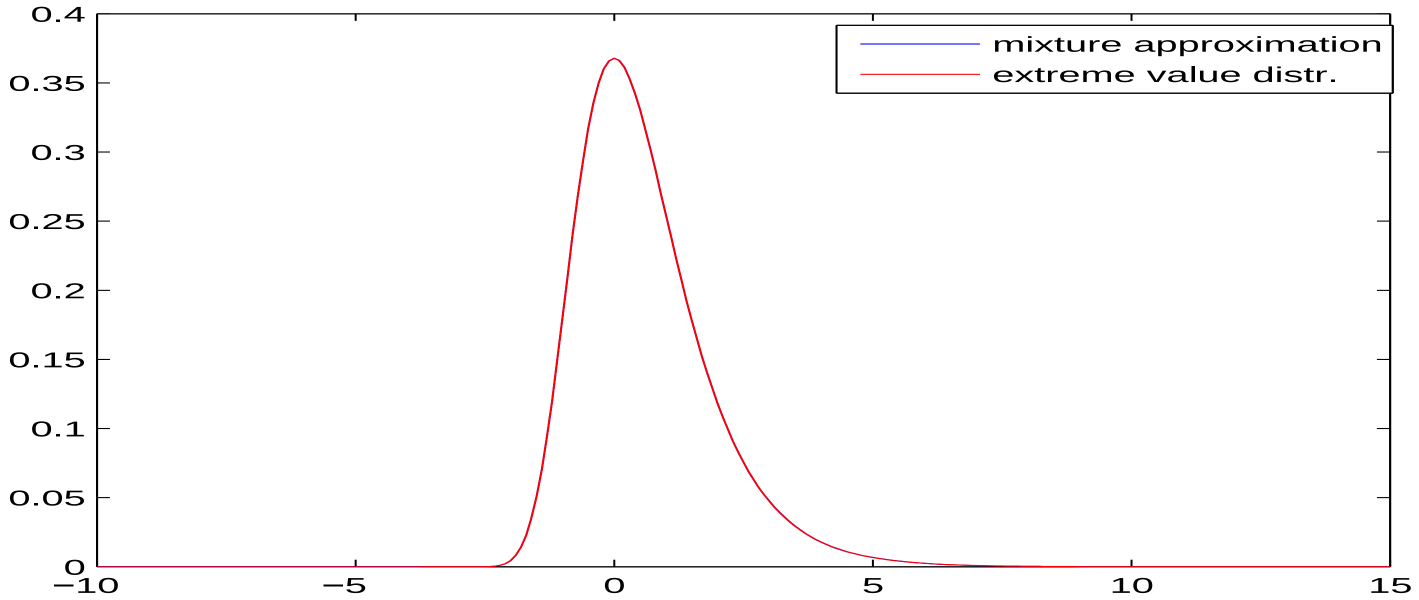
Approximate the density $p(\varepsilon)$ of the type I extreme value distribution by a normal mixture of 10 components with parameters m_r and s_r and weight w_r for the rth component:

$$p(\varepsilon) = \exp\{-\varepsilon - e^{-\varepsilon}\} \approx \sum_{r=1}^{10} w_r f_N(\varepsilon; m_r, s_r^2). \tag{1}$$

The mixture was estimated in Frühwirth-Schnatter and Frühwirth (2007) by minimizing the Kullback-Leibler distance of the estimated mixture from the exact density:

w_r	0.00397	0.0396	0.168	0.147	0.125	0.101	0.104	0.116	0.107	0.088
m_r	5.09	3.29	1.82	1.24	0.764	0.391	0.0431	-0.306	-0.673	-1.06
s_r^2	4.5	2.02	1.1	0.422	0.198	0.107	0.0778	0.0766	0.0947	0.146

Mixture Approximation



The mixture approximation to the type I extreme value distribution

Auxiliary Mixture Sampling for Count Data

- 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:

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

- Substitute the true distribution of ε_{tj} by a mixture of normal distributions and introduce the component indicator r_{tj} as missing data:

$$- \log \tau_{tj} = \log \lambda_t + m_{r_{tj}} + \varepsilon_{tj}, \quad \varepsilon_{tj} \sim \mathcal{N}\left(0, s^2_{r_{tj}}\right).$$

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.

Regression Modelling of Count Data

Consider following regression model:

$$y_t \sim \mathcal{P}\left(\exp(\mathbf{x}_t \boldsymbol{\beta})\right), \tag{2}$$

where $\boldsymbol{\beta}$ is an unknown regression parameter and $\mathbf{x}_t$ is a row vector containing the regressors relevant for that particular regression model, including 1 for the intercept.

Conditional on knowing the auxiliary variables $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_T)$, where $\mathbf{z}_t = (\tau_{tj}, r_{tj}, j = 1, \ldots, y_t + 1)$ a standard linear regression model with heteroscedastic errors results:

$$-\log \tau_{tj} = \mathbf{x}_t \boldsymbol{\beta} + m_{r_{tj}} + \varepsilon_{tj}, \quad \varepsilon_{tj} \sim \mathcal{N}\left(0, s^2_{r_{tj}}\right). \tag{3}$$

Data augmentation and MCMC (Tanner and Wong, 1987): estimating $\boldsymbol{\beta}$ from (2) is equivalent with estimating the augmented vector $(\boldsymbol{\beta}, \mathbf{z})$ from (3).

Auxiliary mixture sampling

(a) Sample $\boldsymbol{\beta}$ from $p(\boldsymbol{\beta}|\mathbf{z},\mathbf{y}) \sim \mathcal{N}\left(\boldsymbol{\alpha}_N, \mathbf{A}_N\right)$-distribution

(b) for each $t = 1, \ldots, T$, sample $\mathbf{z}_t = (\tau_{tj}, r_{tj}, j = 1, \ldots, y_t + 1)$ from $p(\mathbf{r}_t|\boldsymbol{\tau}_t, \boldsymbol{\beta}, \mathbf{y})p(\boldsymbol{\tau}_t|\boldsymbol{\beta}, \mathbf{y})$.

Step (b) is easy to implement:

- the first y_t arrival times $\tau_{t,1}, \ldots, \tau_{t,y_t}$ are distributed as the order statistics of n $\mathcal{U}\left[0,1\right]$-distributed random variables
- final inter-arrival time: $\tau_{t,y_t+1} = (1 - \sum_{j=1}^{y_t} \tau_{tj}) + \mathcal{E}\left(\lambda_t\right)$
- the indicators are sampled as for finite mixture models

Application to Road Safety Data

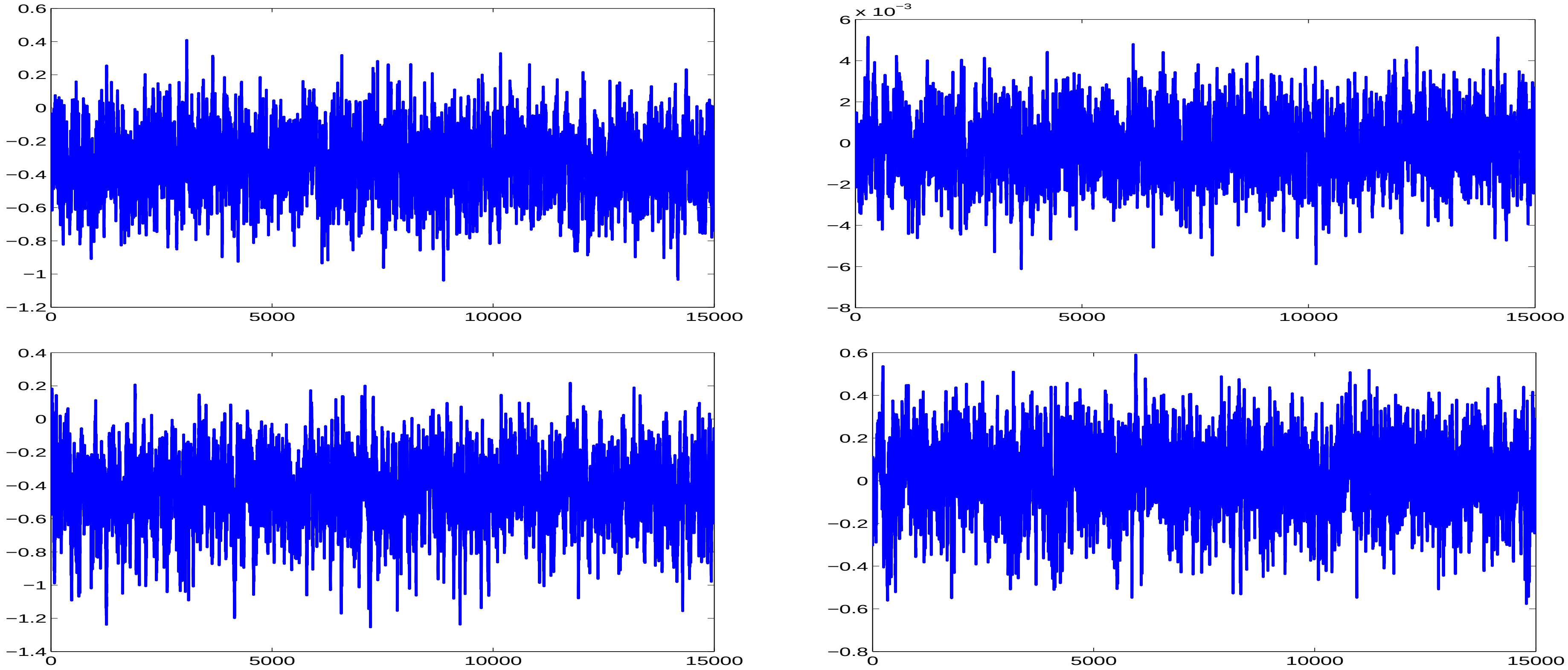


Figure 1: Road safety data: 15 000 draws (burn-in of 5000 removed) for for a model including all effects; top: intervention effect (left), linear trend component (right); bottom: seasonal effects (left: July, right: October)

Estimating Marginal Likelihoods

A direct extension of Chib's estimator (Chib, 1995) of the marginal likelihood $p(\mathbf{y}|\mathcal{M}_k)$ for non-Gaussian fixed parameter models is feasible using the MCMC output of auxiliary mixture sampling for each model (Frühwirth-Schnatter and Wagner, 2007a):

$$\hat{p}_{CH}(\mathbf{y}|\mathcal{M}_k) = \frac{p(\mathbf{y}|\boldsymbol{\vartheta}_k^*)p(\boldsymbol{\vartheta}_k^*)}{p(\boldsymbol{\vartheta}_k^*|\mathbf{y})}, \tag{4}$$

$$\hat{p}(\boldsymbol{\vartheta}_k^*|\mathbf{y}) = \frac{1}{M}\sum_{m=1}^{M} p(\boldsymbol{\vartheta}_k^*|\mathbf{y}, \mathbf{z}^{(m)}).$$

$\boldsymbol{\vartheta}_k$ is the unknown regression parameter in model $\mathcal{M}_k$.

Importance sampling and bridge sampling: importance density constructed from to the MCMC draws obtained by auxiliary mixture sampling

Marginal likelihoods for the Road Safety Data

Table 1: Marginal likelihoods for various Poisson regression models

Model $\mathcal{M}_k$	$\log \hat{p}_{IS}(\mathbf{y}\|\mathcal{M}_k)$	$\log \hat{p}_{BS}(\mathbf{y}\|\mathcal{M}_k)$	$\log \hat{p}_{CH}(\mathbf{y}\|\mathcal{M}_k)$
constant risk	-372.56(0.002)	-372.56(0.003)	-372.55(0.016)
trend	-374.90(0.005)	-374.88(0.005)	-374.93(0.023)
seas	-376.38(0.233)	-374.89(0.091)	-373.46(0.297)
seas, trend	-377.20(0.477)	-377.53(0.096)	-375.56(0.554)
intervention	-368.73(0.007)	-368.73(0.004)	-368.79(0.026)
intervention, seas	-372.40(0.341)	-371.37(0.100)	-368.72(0.349)
intervention, trend	-377.81(0.897)	-378.51(0.105)	-375.21(0.036)
intervention, seas, trend	-378.77(0.446)	-378.56(0.106)	-375.96(0.547)

The Variable Selection Approach

Identify zero regression effects during MCMC estimation (George and McCulloch, 1993, 1997)

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

$$\begin{aligned} &\beta_j = 0, && \text{if } \delta_j = 0, \\ &\beta_j \text{ unrestricted}, && \text{if } \delta_j = 1. \end{aligned}$$

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

$$\begin{aligned} y_t &\sim \mathcal{P}\left(x_{i1}\beta_1 + x_{i4}\beta_4 + \beta_5\right) \\ y_t &\sim \mathcal{P}\left(\delta_1 x_{i1}\beta_1 + \delta_2 x_{i2}\beta_2 + \delta_3 x_{i3}\beta_3 + \delta_4 x_{i4}\beta_4 + \delta_5\beta_5\right) \end{aligned} \tag{5}$$

with $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5)' = (1, 0, 0, 1, 1)'$.

Difficulties with MCMC Estimation

Joint estimation of $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_d)$ and $\boldsymbol{\beta}$ following (George and McCulloch, 1993, 1997):

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

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

It is important to sample the indicators δ_j from the marginal density $p(\delta_j|\boldsymbol{\delta}_{-j}, \mathbf{y})$, where $\boldsymbol{\beta}$ is integrated out.

This is not possible for non-Gaussian models, straightforward only for normal regression models.

Variable selection using Auxiliary mixture sampling

Variable selection for Poisson regression models is possible using auxiliary mixture sampling which is based on introducing the same auxiliary variables $\mathbf{z}$ as before, e.g. for (5):

$$-\log \tau_{tj} = \delta_1 x_{i1}\beta_1 + \delta_2 x_{i2}\beta_2 + \delta_3 x_{i3}\beta_3 + \delta_4 x_{i4}\beta_4 + \delta_5 \beta_5 + m_{r_{tj}} + \varepsilon_{tj},$$

$$\varepsilon_{tj} \sim \mathcal{N}\left(0, s^2_{r_{tj}}\right).$$

1. Sampling of the indicators $\boldsymbol{\delta}$ recursively from $p(\delta_j|\boldsymbol{\delta}_{-j}, \mathbf{z}, \mathbf{y})$, $j = 1, \ldots, d$ (closed form expression);
2. Sampling of the non-zero regression parameters $\boldsymbol{\beta}^{\boldsymbol{\delta}}$ from a normal posterior distribution $p(\boldsymbol{\beta}^{\boldsymbol{\delta}}|\boldsymbol{\delta}, \mathbf{z}, \mathbf{y})$
3. Joint sampling of $\mathbf{z}$ from $p(\mathbf{z}|\mathbf{y}, \boldsymbol{\beta})$

Results for the Children Data

We start with the most general model and include all factors from the beginning.

The variable selection procedure was carried out with 3 indicators for the intervention effect, the linear trend, and the seasonal effect (only one indicator for all seasonal dummies).

All possible 2^3 models are considered to have equal prior probability.

Variable selection was carried out using the normal prior $\mathcal{N}(\log(y_1/e_1), 1) = \mathcal{N}(-9.0084, 1)$ for the intercept and standard normal priors for the other effects

Results are from 20000 iterations of the auxiliary mixture sampler after a burnin of 10000.

Posterior analysis

Posterior probability for the indicators to be non-zero (here $M = 10\,000$):

$$\Pr(\delta_j = 1 | \cdot) = \frac{1}{M} \sum_{m=1}^{M} \delta_j^{(m)},$$

Table 2: Results of variable selection

Parameter	intervention	linear trend	seasonal effect
indicator	0.9841	0.0022	1.0000

Posterior analysis

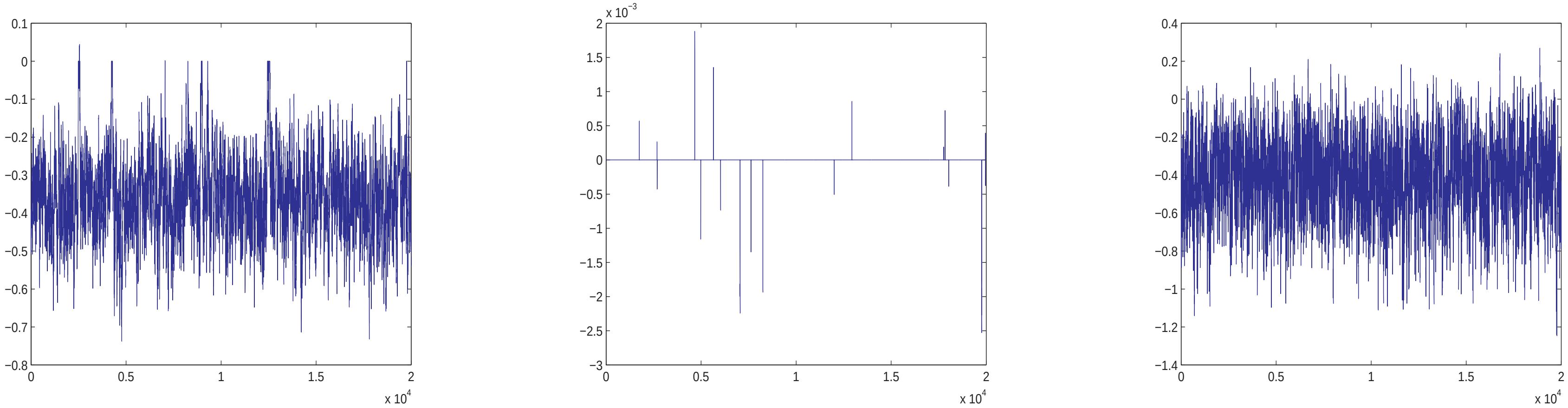


Figure 2: MCMC draws for intervention effect (left hand side), linear trend (middle) and seasonal dummy for July (right hand side)

The Variable Selection Approach

- Is feasible, even if the total number of 2^d possible models is rather large (different prior on the indicators, see e.g. Smith and Kohn (2002))

- But also for a small set of models it turns out to be an attractive alternative to computing marginal likelihoods

- Allows covariate selection for binary data (Holmes and Held, 2006; Tüchler, 2007) and other non-Gaussian data using auxiliary mixture sampling

- Allows covariance selection for Gaussian random effects models (Chen and Dunson, 2003; Frühwirth-Schnatter and Tüchler, 2007) and for binary random effects models using auxiliary mixture sampling (Tüchler, 2007)

- Model selection for Gaussian and non-Gaussian state space models (Frühwirth-Schnatter and Wagner, 2007b)

Model Selection Problem for State Space Models

Time series modelling through the basic structural model for Gaussian as well as non-Gaussian data (binary data, multinomial data, count data):

- Choice of the components in the structural time series model
- Decide, if these components are deterministic or stochastic
- Difficulties with marginal likelihoods
- Extend the variable selection approach to Gaussian state space models
- Use variable selection and auxiliary mixture sampling for non-Gaussian state space models

The Dynamic Linear Trend Model

Introduce three binary indicator δ, γ_1 and γ_2. We introduce separate indicators for the fixed and the really dynamic part:

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2_\varepsilon\right) \\
\mu_t &= \mu_{t-1} + \delta a_0 + \gamma_2 (a_{t-1} - \delta a_0) + \gamma_1 \omega_{1t}, \qquad \omega_{1t} \sim \mathcal{N}\left(0, \theta_1\right), \\
a_t &= a_{t-1} + \omega_{2t}, \qquad \omega_{2t} \sim \mathcal{N}\left(0, \theta_2\right).
\end{aligned}
$$

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

The non-centered parameterization

The following state space model is a noncentered paramerization of the dynamic linear trend model:

$$\tilde{m}_t = \tilde{m}_{t-1} + \tilde{\omega}_{1t}, \qquad \tilde{\omega}_{1t} \sim \mathcal{N}(0,1), \tag{6}$$

$$\tilde{a}_t = \tilde{a}_{t-1} + \tilde{\omega}_{2t}, \qquad \tilde{\omega}_{2t} \sim \mathcal{N}(0,1), \tag{7}$$

$$\tilde{A}_t = \tilde{A}_{t-1} + \tilde{a}_{t-1}, \tag{8}$$

with $\tilde{m}_0 = \tilde{a}_0 = \tilde{A}_0 = 0$. Combine these state equations with following observation equation:

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

where μ_0 and a_0 are the starting value for the level and the drift component and θ_1 and θ_2 are equal to the variances in the dynamic linear trend model.

All 8 combinations of the indicators are identifiable and easy to interpret.

Extension to the basic structural model

Initial seasonal pattern $\mathbf{s}_0 = (s_{-11}, \ldots, s_0)$ with $s_{-11} + \ldots + s_0 = 0$.

Additional binary indicators δ_2 and γ_3 for the fixed and the really dynamic part of the seasonal pattern, where $\mathbf{s}_0 = \mathbf{0}$ if $\delta_2 = 0$ and $\theta_3 = 0$ if $\gamma_3 = 0$:

$$
\begin{aligned}
y_t &= \mu_t + \delta_2 s_{t,0} + \gamma_3 (s_t - \delta_2 s_{t,0}) + \varepsilon_t, \\
\mu_t &= \mu_{t-1} + \delta_1 a_0 + \gamma_2 (a_{t-1} - \delta_1 a_0) + \gamma_1 \omega_{1t}, \\
a_t &= a_{t-1} + \omega_{2t}, \\
s_t &= -s_{t-1} - \cdots - s_{t-11} + \omega_{3t}, \qquad \omega_{3t} \sim \mathcal{N}(0, \theta_3) \qquad (9)
\end{aligned}
$$

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

The non-centered parameterization

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

$$\tilde{s}_t = -\tilde{s}_{t-1} - \cdots - \tilde{s}_{t-S+1} + \tilde{\omega}_{3t}, \qquad \tilde{\omega}_{3t} \sim \mathcal{N}\left(0,1\right), \tag{10}$$

where $\tilde{s}_{-S+1} = \ldots = \tilde{s}_0 = 0$. Combine state equation (10) with the state equations (6) to (8) and following observation equation:

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

$s_{t,0}$ is the jth component of the starting value s_0 of the seasonal component where j is equal to the season corresponding to time t, i.e. $j = 1 + (t-1) \mod S$.

Comments

All 2^5 combinations of the indicators are identifiable and easy to interpret

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

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

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

$$\begin{aligned} y_t \quad = \quad & \mu_0 + \delta_1 t a_0 + \delta_2 s_{t,0} + \\ & + \gamma_1(\pm\sqrt{\theta_1})\tilde{m}_t + \gamma_2(\pm\sqrt{\theta_2})\tilde{A}_t + \gamma_3(\pm\sqrt{\theta_3})\tilde{s}_t + \varepsilon_t. \end{aligned} \tag{11}$$

Choosing Priors

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

Use normal priors for μ_0, a_0, and $\mathbf{s}_0$ (same as in a regression model)

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 centered at 0

This prior is less influential than the inverted Gamma prior if the variance is close to 0 (this is likely to happen in a variable selection environment)

Choosing Priors

Example: local level model

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

Prior on the variance θ_1 may be influential when testing $\theta_1 = 0$ versus $\theta_1 > 0$, if the true value is close to 0; compare:

- the standard conditional conjugate prior, $\theta_1 \sim \mathcal{G}^{-1}\left(c_0, C_0\right)$
- a normal prior on the signed square root, $\pm\sqrt{\theta_1} \sim \mathcal{N}\left(0, 1\right)$

Choosing Priors

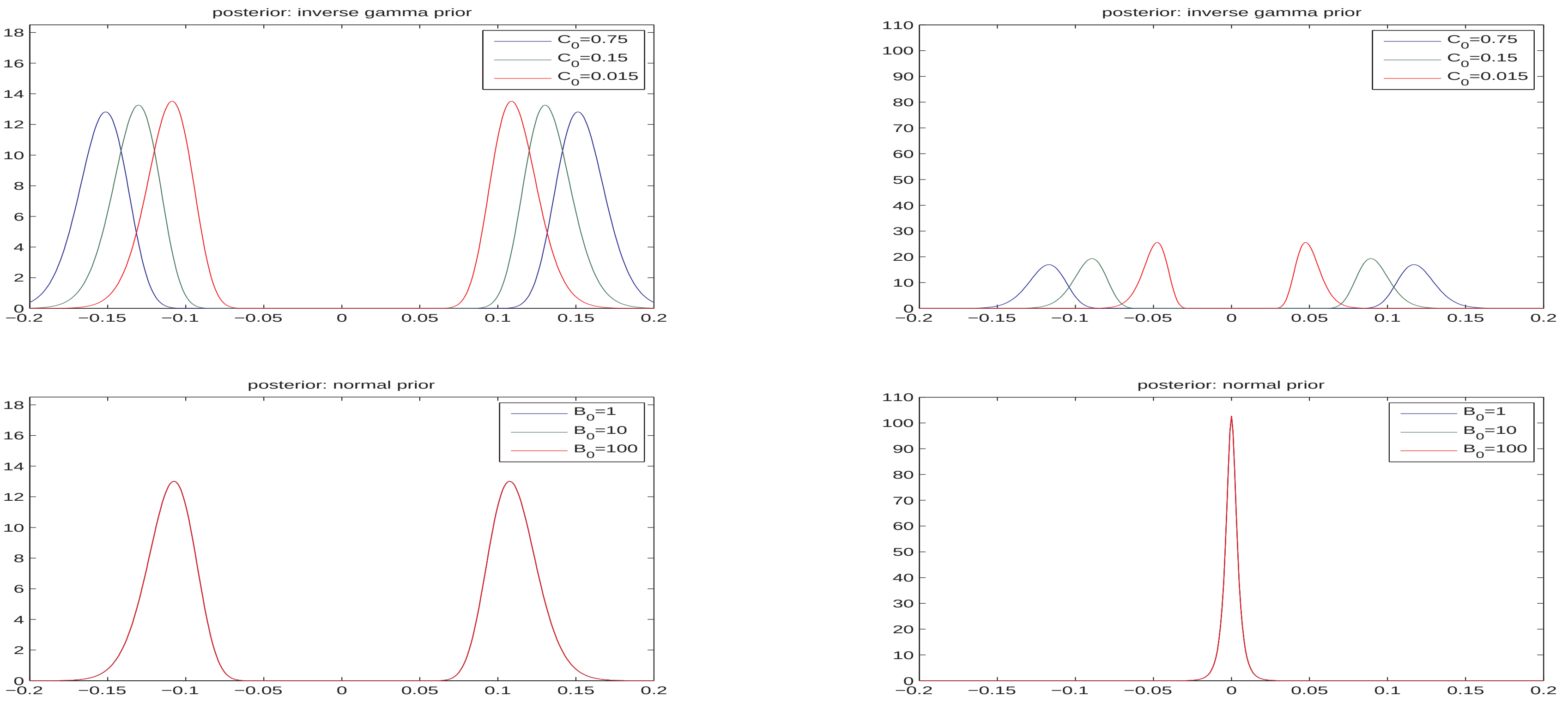


Figure 3: Posterior density for $\pm\sqrt{\theta_1}$ under different priors; top: $\theta_1 \sim \mathcal{G}^{-1}(0.5, C_0)$, bottom: $\pm\sqrt{\theta_1} \sim \mathcal{N}(0, B_0)$; left: dynamic model with $\theta_1 = 0.01$; right: static model with $\theta_1 = 0$; $\sigma_\varepsilon^2 = 1$, $T = 100$

Gibbs sampling scheme

1. Sampling of the indicators $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ recursively from $p(\delta_j|\boldsymbol{\delta}_{-j},\boldsymbol{\gamma},\boldsymbol{\beta},\mathbf{y})$ and $p(\gamma_j|\boldsymbol{\gamma}_{-j},\boldsymbol{\delta},\boldsymbol{\beta},\mathbf{y})$ using the regression model (11) conditional on the state vector $\boldsymbol{\beta}=(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_T)$, $\boldsymbol{\beta}_t=(\tilde{m}_t,\tilde{a}_t,\tilde{A}_t,\tilde{s}_t)$ (closed form expression because the model is conditionally normal given $\boldsymbol{\beta}$);

2. Conditional on $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, sampling of the unrestricted fixed effects μ_0, a_0, $\mathbf{s}_0$ and the unrestricted variances $\pm\sqrt{\theta_j}$ using the regression model (11)

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

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

Model Selection of 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:

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

Application to Road Safety Data

Basic structural model (Harvey and Durbin, 1986):

$$y_t \sim \mathcal{P}\left(e_t \lambda_t\right), \qquad \log \lambda_t = \mu_t + \delta_2 s_{t,0} + \gamma_3 (s_t - \delta_2 s_{t,0}),$$

$$\mu_t = \mu_{t-1} + \delta_1 a_0 + \gamma_2 (a_{t-1} - \delta_1 a_0) + \gamma_1 \omega_{1t}, \qquad \omega_{1t} \sim \mathcal{N}\left(0, \theta_1\right),$$

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

$$s_t = -s_{t-1} - \cdots - s_{t-11} + \omega_{3t}, \qquad \omega_{3t} \sim \mathcal{N}\left(0, \theta_3\right)$$

Modification of trend component for time of intervention, $t = t_{int}$, introduce indicator δ_3 for the intervention effect:

$$\mu_t = \mu_{t-1} + \delta_1 a_0 + \gamma_2 (a_{t-1} - \delta_1 a_0) + \delta_3 \Delta + \gamma_1 \omega_{1t}.$$

Auxiliary mixture sampling and variable selection

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

$$
\begin{aligned}
-\log \tau_{tj} - \log e_t &= \mu_t + \delta_2 s_{t,0} + \gamma_3 (s_t - \delta_2 s_{t,0}) + m_{r_{tj}} + \varepsilon_t, \\
\varepsilon_t &\sim \mathcal{N}\left(0, s^2_{r_{tj}}\right). \qquad (12)
\end{aligned}
$$

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 (12) (conditional on the auxiliary variables $\mathbf{z}$)

2. Sample the auxiliary variables $\mathbf{z}$ (same procedure as above given the current values of $\log \lambda_t, t = 1, \ldots, T$).

Time Series of Road Accidents of Children

Priors for the initial values:

$$\mu_0 \sim \mathcal{N}\left(\log(y_1/e_1), 1\right), \quad a_0 \sim \mathcal{N}\left(0, 1\right), \quad \Delta \sim \mathcal{N}\left(0, 1\right),$$
$$(s_{-1}, \ldots, s_{-11})' \sim \mathcal{N}_{11}\left(\mathbf{0}, \mathbf{I}\right).$$

Priors for the variances:

$$\pm\sqrt{\theta_1} \sim \mathcal{N}\left(0, 1\right), \quad \pm\sqrt{\theta_2} \sim \mathcal{N}\left(0, 1\right), \quad \pm\sqrt{\theta_3} \sim \mathcal{N}\left(0, 1\right).$$

All 2^6 models have same prior probabilities

MCMC sampling was carried out for 20000 iterations after a burn-in of 10000.

Time Series of Road Accidents of Children

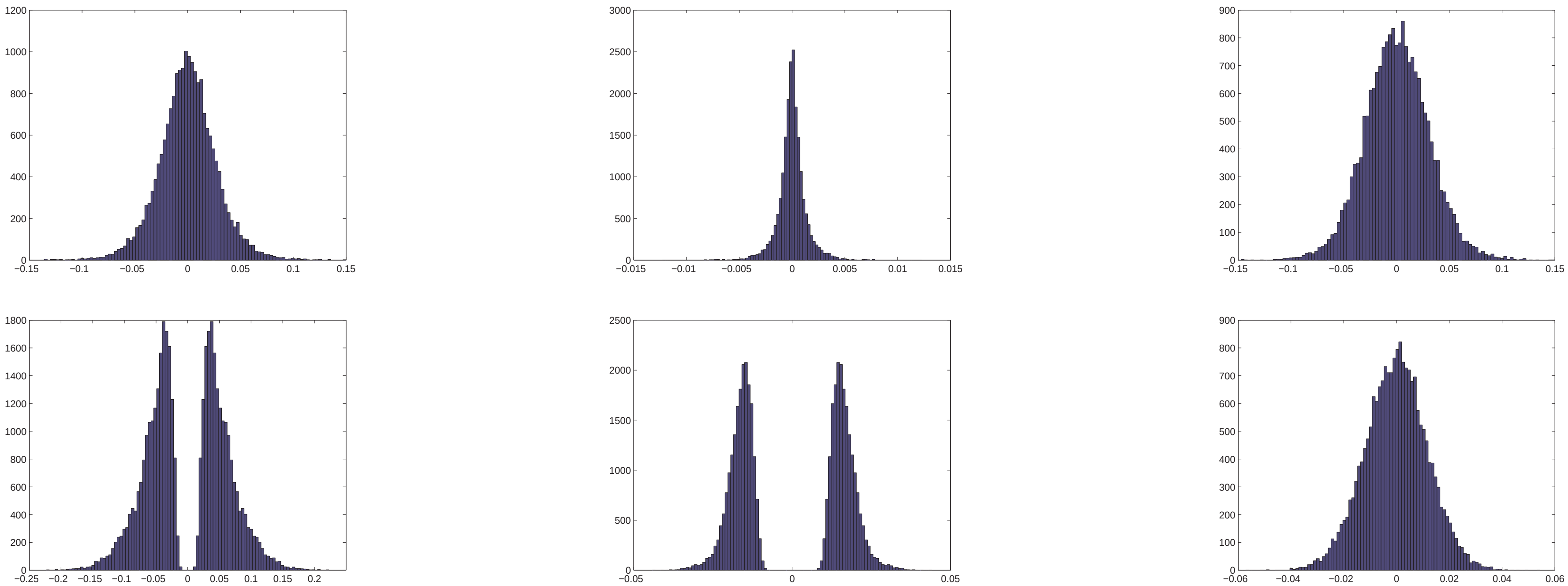


Figure 4: Histograms for $\pm\sqrt{\theta_1}$ (left), $\pm\sqrt{\theta_2}$ (middle) and $\pm\sqrt{\theta_3}$ (right); bottom: $\mathcal{G}^{-1}(0.1, 0.001)$-prior for θ_1 and θ_2, $\mathcal{N}(0,1)$ prior for $\pm\sqrt{\theta_3}$ (Frühwirth-Schnatter and Wagner, 2006a)

MCMC draws for $\delta = 1$ and $\gamma = 1$

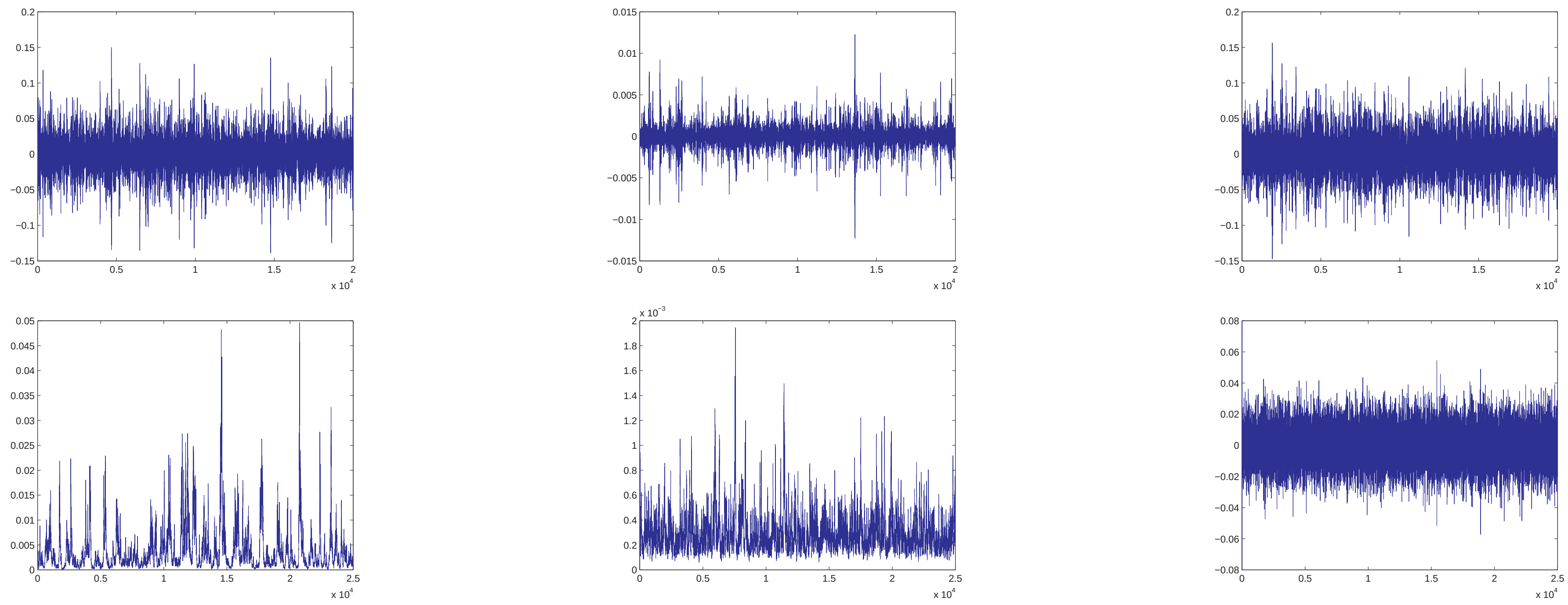


Figure 5: MCMC draws for $\pm\sqrt{\theta_1}$ (left), $\pm\sqrt{\theta_2}$ (middle) and $\pm\sqrt{\theta_3}$ (right); bottom: θ_1 (left), θ_2 (middle) and $\pm\sqrt{\theta_3}$ (right) (Frühwirth-Schnatter and Wagner, 2006a)

Exploratory Bayesian Analysis with $\delta = 1$ and $\gamma = 1$

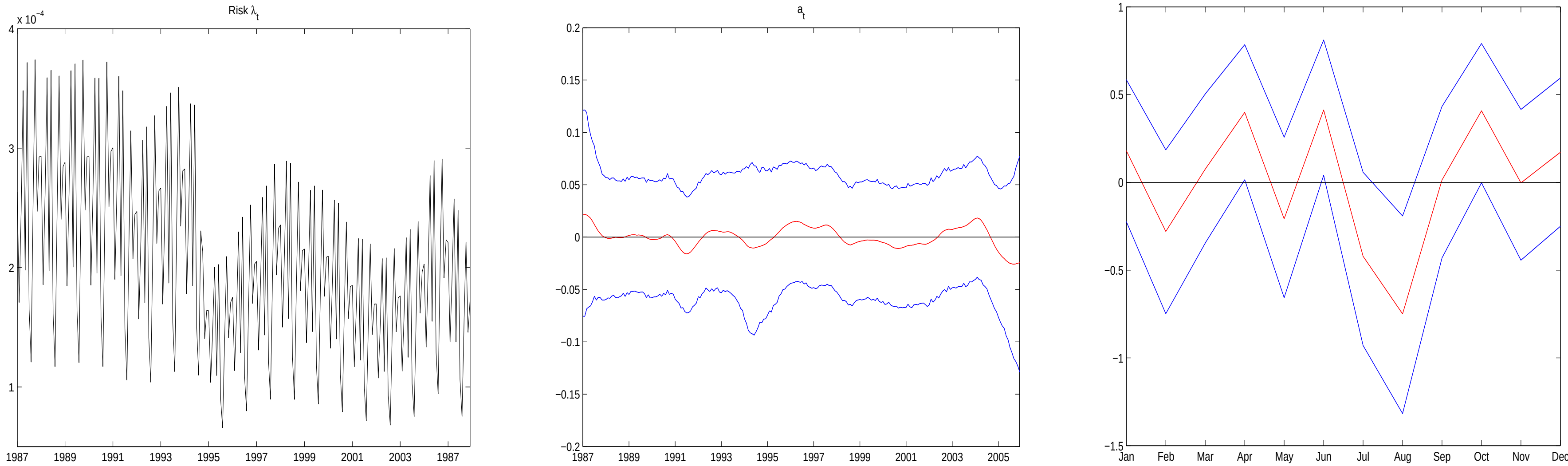


Figure 6: Estimated risk λ_t (left), drift a_t (middle), seasonal components s_t in year 2005 (right)

Results of Variable Selection

Table 3: Results of variable selection

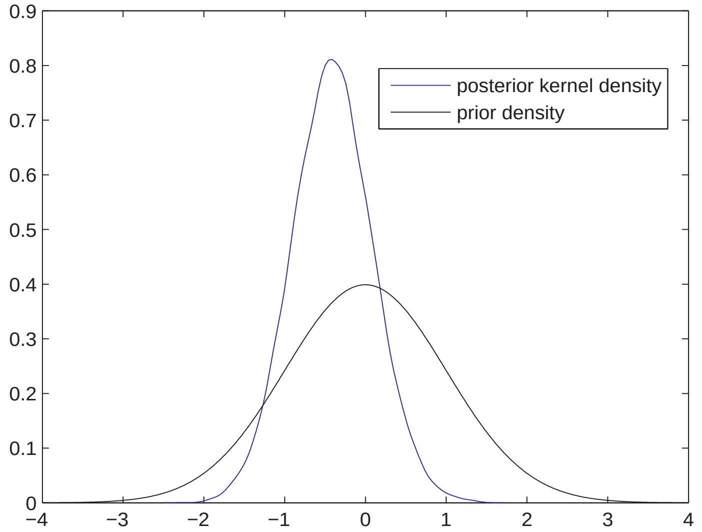
	fixed effects			process variances		
Parameter	intervention	a_0	s_0	γ_1	γ_2	γ_3
indicator	0.9705	0.0027	1.0000	0.0185	0.0005	0.0246

Simple Poisson regression model

- with seasonal pattern
- intervention effect significant

Results of Variable Selection

Gain of statistical efficiency for the parameter of interest



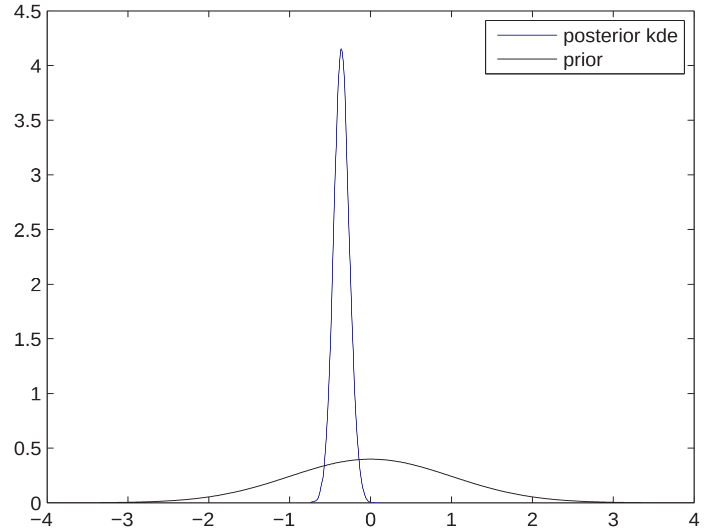


Figure 7: posterior density of the intervention effect Δ in comparison to the prior; left: unrestricted basic structural model, right: Poisson regression model with seasonal pattern (selected model)

Marginal Likelihoods for Non-Gaussian State Space Models

As for fixed parameter models, Chib's estimator is based on identity (4):

$$\hat{p}_{CH}(\mathbf{y}|\mathcal{M}_k) = \frac{p(\mathbf{y}|\boldsymbol{\vartheta}_k^*)p(\boldsymbol{\vartheta}_k^*)}{\hat{p}(\boldsymbol{\vartheta}_k^*|\mathbf{y})}, \tag{13}$$

where $\boldsymbol{\vartheta}_k$ contains all unconstrained variances, initial values $a_0, \mu_0, s_{-1}, \ldots, s_{-11}$ and Δ.

If a numerical value for the integrated likelihood function $p(\mathbf{y}|\boldsymbol{\vartheta}_k^*)$ is available, then (13) could be implemented as before by estimating the posterior ordinate $\hat{p}(\boldsymbol{\vartheta}_k^*|\mathbf{y})$ by means of the draws obtained from auxiliary mixture sampling:

$$\hat{p}(\boldsymbol{\vartheta}_k^*|\mathbf{y}) = \frac{1}{M}\sum_{m=1}^{M} p(\boldsymbol{\vartheta}_k^*|\boldsymbol{\beta}_k^{(m)}, \mathbf{z}_k^{(m)}, \mathbf{y}). \tag{14}$$

Use particle filtering to obtain $p(\mathbf{y}|\boldsymbol{\vartheta}_k^*)$.

Marginal Likelihoods for Non-Gaussian State Space Models

Complete-likelihood estimator avoids particle filtering:

$$\hat{p}_{CDL}(\mathbf{y}|\mathcal{M}_k) = \frac{p(\mathbf{y}|\boldsymbol{\vartheta}_k^*, \boldsymbol{\beta}_k^*)\, p(\boldsymbol{\beta}_k^*|\boldsymbol{\vartheta}_k^*) p(\boldsymbol{\vartheta}_k^*)}{\hat{p}(\boldsymbol{\vartheta}_k^*, \boldsymbol{\beta}_k^*|\mathbf{y})}, \tag{15}$$

where $\boldsymbol{\beta}_k$ is the state vector. The complete-data likleihood $p(\mathbf{y}|\boldsymbol{\vartheta}_k^*, \boldsymbol{\beta}_k^*)$ is known analytically.

The joint posterior ordinate $\hat{p}(\boldsymbol{\vartheta}_k^*, \boldsymbol{\beta}_k^*|\mathbf{y})$ is estimated by means of the draws obtained from auxiliary mixture sampling:

$$\hat{p}(\boldsymbol{\vartheta}_k^*, \boldsymbol{\beta}_k^*|\mathbf{y}) = \frac{1}{M} \sum_{m=1}^{M} p(\boldsymbol{\vartheta}_k^*, \boldsymbol{\beta}_k^*|\mathbf{z}_k^{(m)}, \mathbf{y}). \tag{16}$$

Marginal Likelihoods for Non-Gaussian State Space Models

Table 4: Marginal likelihoods for the road Safety Data

Model $\mathcal{M}_k$	$\log \hat{p}_{CH}(\mathbf{y}\|\mathcal{M}_k)$	$\log \hat{p}_{CDL}(\mathbf{y}\|\mathcal{M}_k)$
Poisson reg. – seas., int.	-368.7187 (0.3490)	—
loc. level – fixed seas., no int.	-377.7182 (0.4071)	-376.3957 (0.6514)
loc. level – fixed seas., int.	-378.6764 (0.6355)	-379.1788 (0.9049)
loc. trend – fixed seas., no int.	-414.1827 (0.3510)	-404.9596 ($9.0568.10^{28}$)
loc. trend – fixed seas., int.	-415.7098 (0.2607)	-403.0040($3.6730.10^{30}$)
basic struct. model, no int.	-418.9280 (0.4512)	-405.8012($2.3599.10^{27}$)
basic struct. model, int.	-420.3245 (0.5653)	-407.4017($8.9277.10^{30}$)

Improved Auxiliary mixture sampling for Poisson data

Frühwirth-Schnatter, Frühwirth, Held, and Rue (2007): reduce the dimension of the auxiliary variables $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_T)$

- Original sampler: $\mathbf{z}_t = \{(\tau_{tj}, r_{tj}), j = 1, \ldots, y_t + 1\}$ $\Rightarrow$ dimension of $\mathbf{z}_t$ is equal to $2(y_t + 1)$ and increases with y_t

- Improved sampler: for all observations y_t with $y_t > 0$: $\mathbf{z}_t = (\tau^{\star}_{t1}, \tau^{\star}_{t2}, r_{t1}, r_{t2})$ $\Rightarrow$ dimension of $\mathbf{z}_t$ is equal to 4 regardless of y_t

Achieved by considering only the first inter arrival time after 1 and the last arrival times before 1.

Improved Auxiliary mixture sampling for Poisson data

Therefore

$$\tau_{t1}^{\star} = \tau_{t,y_t+1} \sim \log \mathcal{E}\left(\lambda_t\right),$$
$$\tau_{t2}^{\star} = \sum_{t=1}^{y_t} \tau_{t,j} \sim \mathcal{G}\left(y_t, \lambda_t\right),$$

and

$$-\log \tau_{t1}^{\star} = \log \lambda_t + \varepsilon_{t1}, \qquad \varepsilon_{t1} \sim -\log \mathcal{E}\left(1\right),$$
$$-\log \tau_{t2}^{\star} = \log \lambda_t + \varepsilon_{t2}, \qquad \varepsilon_{t2} \sim -\log \mathcal{G}\left(y_t, 1\right).$$

For each integer value $n = 1, \ldots$ a mixture approximation with decreasing number of components has been developed to approximate the distribution of ε_{t2}.

Considerable reduction in computing time for medium to large counts

Summary

Bayesian approach offers in principle a coherent way of model selection, practical Bayesian model selection, however, is a numerical challenge

Variable selection approach offers a lot of advantages compared to estimating the marginal likelihood for each model

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

Prior choices may be influential

Summary

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

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

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

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and

Summary

Markov switching models using bridge sampling techniques. The Econometrics Journal 7, 143–167.

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

Frühwirth-Schnatter, S., R. Frühwirth, L. Held, and H. Rue (2007). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. IFAS Research Report 2007-25, `http://ifas.jku.at`, Johannes Kepler University Linz.

Frühwirth-Schnatter, S. and R. Tüchler (2007). Bayesian parsimonious covariance estimation for hierarchical linear mixed models. Statistics and Computing. to appear.

Frühwirth-Schnatter, S. and H. Wagner (2006a). 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 H. Wagner (2006b). Data augmentation and Gibbs

Summary

sampling for regression models of small counts. Student 5, 221–234. Available at http://www.ifas.jku.at/ as Research Report IFAS 2004-04.

Frühwirth-Schnatter, S. and H. Wagner (2007a). Marginal likelihoods for non-gaussian models using auxiliary mixture sampling. Research Report IFAS 2007-24, http://www.ifas.jku.at/.

Frühwirth-Schnatter, S. and H. Wagner (2007b). Stochastic model specification search for gaussian and non-gaussian state space models. Research Report IFAS 2007-XX, http://www.ifas.jku.at/, in preparation.

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

George, E. I. and R. McCulloch (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. and J. Durbin (1986). The effects of seat belt legislation on british road casualties: A case study in structural time series modelling. Journal of

Summary

the Royal Statistical Society, Ser. A 149, 187–227. CHECK.

Holmes, C. C. and L. Held (2006). Bayesian auxiliary variable models for binary and multinomial regression. Bayesian Analysis 1, 145–168.

Lindley, D. V. (1957). A statistical paradoxon. Biometrika 44, 187–192. CHECK.

Meng, X.-L. and W. H. Wong (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. Statistica Sinica 6, 831–860.

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

Smith, M. and R. Kohn (2002). Parsimonious covariance matrix estimation for longitudinal data. Journal of the American Statistical Association 97, 1141–1153.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–540.

Tüchler, R. (2007). Bayesian variable selection for logistic models using auxiliary mixture sampling. Journal of Computational and Graphical Statistics ADD,

Summary

ADD. to appear.

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