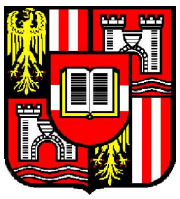

Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2005-11

Auxiliary Mixture Sampling for Parameter-driven Models of Time Series of Counts with Applications to State Space Modelling

Sylvia Frühwirth-Schnatter and Helga Wagner

December 2005

Abstract

In this article we consider parameter-driven models of time series of counts, where the observations are assumed to arise from a Poisson distribution with a mean changing over time according to a latent process. Estimation of these models is carried out within a Bayesian framework using data augmentation and Markov chain Monte Carlo methods. We suggest a new auxiliary mixture sampler, which possesses a Gibbsian transition kernel, where we draw from full conditional distributions belonging to standard distribution families only. Emphasis lies on application to state space modelling of time series of counts. Nevertheless we show that auxiliary mixture sampling is more general than that and may be applied to a wide range of parameter-driven models, including random-effects models and panel data models based on the Poisson distribution.

Key words: count data, data augmentation, finite mixture approximation, Gibbs sampling, partially Gaussian state space models

1 Introduction

Applied statisticians commonly have to deal with time series of counts, modelling the number of events occurring in a given interval. Typical examples are the number of road accidents recorded during a given period or data on disease occurrences. Such data are necessarily non-negative integers and it is often appropriate to assume that the observed process y_t follows a Poisson distribution. To capture the effect of exogenous variables, summarized in the row vector Z_t, a log-linear model could be applied, where

$$y_t|\lambda_t \sim \text{Poisson}\left(\lambda_t\right), \lambda_t = \exp(Z_t^{'}\beta),$$

with λ_t being the mean of the time series observation y_t given β, and β being a vector of unknown coefficients to be estimated from the data. In the standard log-linear model it is assumed that the count observations are independent. To account for the dependence likely to be present in time series data of counts, various extensions of the log-linear model have been suggested which, following Cox (1981), may be broadly characterized as parameter-driven and observation-driven models. For observation driven models, the mean structure of the conditional distribution of y_t given past observations $y_{t-1}, y_{t-2}, \ldots$, is directly specified as a function of these observations, see for instance Zeger and Qaqish (1988). In this article we consider parameter-driven models, where dependence among observations is introduced indirectly through a latent process, for instance a hidden Markov chain as in Leroux and Puterman (1992), or a latent stationary autoregressive process as in Zeger (1988) and Chan and Ledolter (1995). More general state-space models based on a first order hidden Markov process have been considered by, among others, West et al. (1985), and Harvey and Fernandes (1989). A key property of parameter-driven models is that the distribution of y_t is allowed to depend on this latent process, and although the observations are correlated marginally, conditional upon knowing the latent process typically they are independent.

Estimation of parameter-driven Poisson time series models often turns out to be a challenging problem. In some cases, like hidden Markov chain models, maximum

likelihood estimation is straightforward, for other models maximum likelihood is hampered by the fact that the marginal likelihood, where the latent process is integrated out, is not available in closed form. Each evaluation of the likelihood function requires to use some numerical method for solving the necessary high-dimensional integration. One particular useful method in this respect is importance sampling which was applied in Durbin and Koopman (2000) to state space modelling of count data, see also Durbin and Koopman (2001).

Alternatively, estimation of these models is feasible within a Bayesian framework using data augmentation as in Tanner and Wong (1987) and Markov chain Monte Carlo methods, as illustrated first by Zeger and Karim (1991) for generalized linear models with random effects. Since this seminal paper, a number of authors have contributed to Markov chain Monte Carlo estimation of parameter-driven models for count data. We mention here in particular Albert (1992) for Poisson random-effects models, Wakefield et al. (1994) for more general random effect models, Shephard and Pitt (1997) for non-Gaussian time series models based on distributions from the exponential family, Gamerman (1998) for dynamic generalized linear models, Chib et al. (1998) for panel count data models with multiple random effects, Lenk and DeSarbo (2000) for mixtures of generalized linear models with random effects, and Chib and Winkelmann (2001) for correlated multivariate count data. A major difficulties with any of the existing Markov chain Monte Carlo approaches, however, is that practical implementation requires the use of a Metropolis-Hastings algorithm at least for part of the unknown parameter vector, which in turn makes it necessary to define suitable proposal densities in rather high-dimensional parameter spaces. Single-move sampling for this type of models is likely to be very inefficient, see e.g. Shephard and Pitt (1997).

The main contribution of the present article is to show how to design an approximate, yet very accurate, straightforward Gibbs sampling scheme for all unknown quantities, requiring only random draws from standard distributions such as multivariate normals, inverse Gamma, exponential and discrete distributions with a few categories. Although we focus on state space models for Poisson counts, any other model with Poisson counts and some linear structure in the log intensity may estimated in a similar way.

This rather unexpected result is achieved by introducing two sequences of latent variables through data augmentation. The first of these sequences are the unobserved inter-arrival times of suitably chosen Poisson processes. The introduction of this first sequence eliminates the non-linearity of the observation equation, whereas the non-normality of the error term, which is minus the logarithm of a random variable from the standard exponential distribution, remains. The distribution of the error term is then approximated by a mixture of normal distributions in a similar way as in Kim et al. (1998) and Chib et al. (2002) who used a normal mixture approximation for the distribution of the logarithm of a random variable following a χ^2-distribution in the context of stochastic volatility models. By introducing the component indicator of the normal mixture approximation as a second sequence of missing data, the resulting model may be thought of as a partially Gaussian model as in Shephard (1994), and Gibbs sampling becomes feasible. This will be shown to be particularly useful for state space models for Poisson time series, as multi-move-sampling of the whole state process through forward-filtering backward sampling as

in Frühwirth-Schnatter (1994b), Carter and Kohn (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002) is feasible.

The rest of the paper is organized as follows. In Section 2, we introduce in detail our new method of data augmentation for parameter-driven models based on the Poisson distribution, that will be exploited in Section 3 to implement a Gibbs sampling scheme for these models. Applications to state space modelling of Austrian road safety data are considered in Section 4, whereas Section 5 concludes.

2 Data Augmentation for Parameter-Driven Models based on the Poisson Distribution

2.1 Model Specification

Let $y_1, \ldots, y_T$ be a sequence of count data, observed at discrete, evenly spaced time points. In what follows, we assume that $y_t|\lambda_t$ follows a Poisson (λ_t) distribution, where the risk λ_t is allowed to depend on exogenous information $Z_t = (Z_t^1 Z_t^2)$ through fixed model parameters α and time-varying model parameters β_t in the following way:

$$y_t|\lambda_t \sim \text{Poisson}\,(\lambda_t)\,, \tag{1}$$

$$\lambda_t = \exp(Z_t^1\alpha + Z_t^2\beta_t). \tag{2}$$

The precise model for β_t will be left unspecified at this stage, we only assume that the joint distribution $p(\alpha, \beta_1, \ldots, \beta_T|\theta)$ follows a normal distribution, which is allowed to be indexed by an unknown model parameter θ. Furthermore we assume that conditional on knowing $\alpha, \beta_1, \ldots, \beta_T$, the observations $y_t|\lambda_t$ and $y_s|\lambda_s$ are mutually independent.

These model assumptions are sufficient to derive the conditional posterior density $p(\alpha, \beta_1, \ldots, \beta_T|\theta, y)$ formally by Bayes' theorem, given the whole time series $y = (y_1, \ldots, y_T)$, the resulting posterior density, however, in general does not belong to a density from a well-known distribution family. Although $\log \lambda_t$ in (2) is linear in the unknown model parameters $\alpha, \beta_1, \ldots, \beta_T$, the presence of the Poisson distribution in the observation equation (1) causes non-normality as well as non-linearity of the mean λ_t in $\alpha, \beta_1, \ldots, \beta_T$. We are going to demonstrate in this section, how the introduction of two sequences of artificially missing data within a data augmentation scheme eliminates both non-normality and non-linearity and leads to a conditional posterior distribution for $\alpha, \beta_1, \ldots, \beta_T$ that is a multivariate normal distribution, once we conditioned on the artificially missing data.

2.2 Step 1: Data augmentation through hidden inter-arrival times

For each t, the distribution of $y_t|\lambda_t$ may be regarded as the distribution of the number of jumps of an unobserved Poisson process with intensity λ_t, having occurred in the time interval [0,1]. The first step of data augmentation creates such a Poisson process for each observation y_t, $t = 1, \ldots, T$, and introduces the inter-arrival times

$\tau_{tj},\ j = 1, \ldots, (y_t + 1)$ of this Poisson process as missing data. From the basic properties of a Poisson process, the inter-arrival times τ_{tj} are known to follow the Exponential (λ_t)-distribution, therefore:

$$\tau_{tj}|\alpha, \beta_t = \frac{\xi_{tj}}{\lambda_t}, \quad \xi_{tj} \sim \text{Exponential}\,(1)\,.$$

This may be reformulated as following linear model:

$$-\log \tau_{tj}|\alpha, \beta_t = Z_t^1\alpha + Z_t^2\beta_t + \varepsilon_{tj}, \tag{3}$$

where $\varepsilon_{tj} = -\log \xi_{tj}$ with $\xi_{tj} \sim \text{Exponential}\,(1)$.

Let $\tau = \{\tau_{tj}, j = 1, \ldots, (y_t + 1), t = 1, \ldots, T\}$ denote the collection of all inter-arrival times. Our first data augmentation step introduces the inter-arrival times τ as missing data, with two effects. First, the full-conditional posterior distribution $p(\alpha, \beta_1, \ldots, \beta_T|\theta, \tau, y)$ of $\alpha, \beta_1, \ldots, \beta_T$, where additionally to θ and y the inter-arrival times τ appear as conditioning argument, is independent of y: $p(\alpha, \beta_1, \ldots, \beta_T|\theta, \tau, y) = p(\alpha, \beta_1, \ldots, \beta_T|\theta, \tau)$. Second, conditional on τ, we are dealing with model (3), which is non-normal, but where the mean of the observation equation is linear in the unknown model parameters $\alpha, \beta_1, \ldots, \beta_T$:

$$\text{E}(-\log \tau_{tj}|\alpha, \beta_t) = Z_t^1\alpha + Z_t^2\beta_t + 0.57722.$$

2.3 Step 2: Data augmentation through a Mixture Approximation

Whereas the first augmentation steps eliminates the non-linearity of the observation equation, the non-normality of the error term, however, remains. It is important to realize that the error term in (3) may be represented as minus log of a random variable from the standard exponential distribution. The density $p_\varepsilon(\varepsilon)$ of such a random variable is independent of any unknown model parameter and reads:

$$p_\varepsilon(\varepsilon) = \exp\{-\varepsilon - e^{-\varepsilon}\}.$$

To obtain a model that is conditionally Gaussian, we approximate this non-normal density by a normal mixture of R components with parameters m_r and s_r for the r-th component:

$$p_\varepsilon(\varepsilon) = \exp\{-\varepsilon - e^{-\varepsilon}\} \approx q_{R,\varepsilon}(\varepsilon) = \sum_{r=1}^{R} w_r f_{\text{N}}(\varepsilon; m_r, s_r^2). \tag{4}$$

This idea is influenced by the related articles of Kim et al. (1998) and Chib et al. (2002), who used a normal mixture approximation of the density of a $\log \chi^2$-distribution in the context of stochastic volatility models. The appropriate parameters $(w_r, m_r, s_r^2), r = 1, \ldots, R$, however, are different for our problem. They were determined for $R = 2, \ldots, 10$ numerically by minimizing the Kullback-Leibler distance between the true density and the mixture approximation, see Frühwirth-Schnatter and Frühwirth (2005) for further details. The parameters (w_r, m_r, s_r^2) are tabulated in Table 1 for $R = 10$, whereas Figure 1 compares the true density with a normal mixture

Table 1: Normal mixture approximation with 10 components of the density of $-\log \xi$, where $\xi \sim$ Exponential (1)

r	1	2	3	4	5	6	7	8	9	10
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

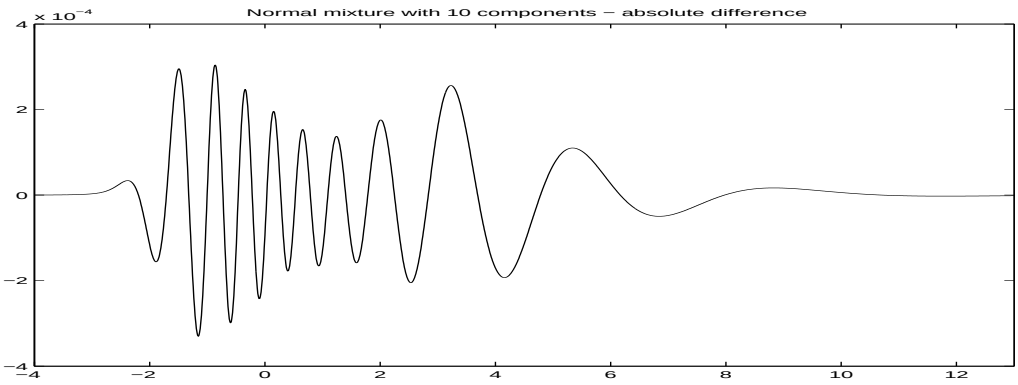

Figure 1: Difference between the ten component normal mixture approximation given by Table 1 and the density of $-\log \xi$, where $\xi \sim$ Exponential (1)

approximation based on 10 components, by plotting the difference between these densities.

Following Kim et al. (1998) and Chib et al. (2002), the density $p_\varepsilon(\varepsilon_{tj})$ in (3) is approximated for each t and j by the mixture approximation $q_{R,\varepsilon}(\varepsilon_{tj})$. The second step of our data augmentation scheme introduces for each ε_{tj} the latent component indicator r_{tj} as missing data. Let $S = \{r_{tj}, j = 1, \ldots, (y_t + 1), t = 1, \ldots, T\}$ denote the collection of all component indicators r_{tj}. The introduction of S as additional missing data has the desirable effect, that conditional on τ and S the non-normal, non-linear model (1) and (2) reduces to a linear, Gaussian model where the mean of the observation equation is linear in the unknown model parameters $\alpha, \beta_1, \ldots, \beta_T$ and the error term follows a normal distribution:

$$-\log \tau_{tj} | \alpha, \beta_t, r_{tj} = Z_t^1 \alpha + Z_t^2 \beta_t + m_{r_{tj}} + \varepsilon_{tj}, \quad \varepsilon_{tj} | r_{tj} \sim \text{Normal}\left(0, s_{r_{tj}}^2\right).$$

Consequently, the conditional posterior $p(\alpha, \beta_1, \ldots, \beta_T | \theta, \tau, S, y)$ is given by

$$\begin{aligned} &p(\alpha, \beta_1, \ldots, \beta_T | \theta, \tau, S, y) \propto \\ &\qquad p(\alpha, \beta_1, \ldots, \beta_T | \theta) \prod_{t=1}^{T} \prod_{j=1}^{y_t+1} f_{\mathrm{N}}(\varepsilon_{tj}; -\log \tau_{tj} - m_{r_{tj}} - Z_t^1 \alpha - Z_t^2 \beta_t, s_{r_{tj}}^2), \end{aligned} \tag{5}$$

which is proportional to a multivariate normal density. This result motivated the sampling scheme that will be described in Section 3.

3 Auxiliary Mixture Sampling for Parameter-driven Models for Time Series of Counts

As mentioned in the introduction, Markov chain Monte Carlo estimation of parameter-driven models for time series of counts which are based on the Metropolis-Hastings algorithm have been considered by many authors, in particular by Zeger and Karim (1991), Albert (1992), Shephard and Pitt (1997), Chib et al. (1998) and Chib and Winkelmann (2001). In Section 2, we were able to show that any parameter-driven model based on the Poisson distribution may be closely approximated by a partially Gaussian model in the sense of Shephard (1994). This very useful result will be exploited in this section to implement auxiliary mixture sampling for rather general parameter-driven models for time series of counts.

3.1 The Basic Three-block Auxiliary Mixture Sampler

A three-block Gibbsian sampler results, if data augmentation as described in the previous section is applied for all observations, by introducing the inter-arrival times $\tau = \{\tau_{tj}, j = 1, \ldots, y_t + 1, t = 1, \ldots, T\}$ and the component indicators $S = \{r_{tj}, j = 1, \ldots, y_t + 1, t = 1, \ldots, T\}$ as missing data. Select a starting value for τ, S, and the unknown model parameter θ, and repeat the following steps:

(a) Multi-move sampling of α and the *whole* sequence $\beta = \{\beta_1, \ldots, \beta_T\}$ from the multivariate normal distribution (5), conditional on τ, S, θ and y;

(b) sample θ conditional on α, β, τ, S, and y;

(c) sample the inter-arrival times τ and the component indicators S conditional on y, θ, α and β by running the following steps for $t = 1, \ldots, T$:

(c1) sample the inter-arrival times $\{\tau_{tj}, j = 1, \ldots, y_t + 1\}$. If $y_t > 0$, sample the order statistics $u_{t,(1)}, \ldots, u_{t,(n)}$ of $n = y_t$ uniformly distributed random variables, see e.g. Robert and Casella (1999, p.47) for details, and define the inter-arrival times τ_{tj} as their increments: $\tau_{tj} = u_{t,(j)} - u_{t,(j-1)}, j = 1, \ldots, n$, where $u_{i,(0)} := 0$. Sample the final arrival time as $\tau_{t,n+1} = 1 - \sum_{j=1}^{n} \tau_{tj} + \xi_t$, where $\xi_t \sim$ Exponential (λ_t);

(c2) sample the component indicators r_{tj} for $j = 1, \ldots, y_t + 1$.

The first two steps are model dependent, but for many models involve only standard draws, as we are dealing with a Gaussian model, once we conditioned on τ and S. Step (c), however, deserves detailed investigation. This step is based on decomposing the joint posterior $p(\tau, S|\theta, \alpha, \beta, y)$ as:

$$p(\tau, S|y, \theta, \alpha, \beta) = p(S|\tau, y, \theta, \alpha, \beta)p(\tau|y, \theta, \alpha, \beta)$$

We first sample the inter-arrival times τ from $p(\tau|y, \theta, \alpha, \beta)$. The inter-arrival times $\{\tau_{tj}, j = 1, \ldots, y_t + 1\}$ are independent for different time points t, given β, θ, α and y:

$$p(\tau|y, \theta, \alpha, \beta) = \prod_{t=1}^{T} p(\tau_{t1}, \ldots, \tau_{t,y_t+1}|y_t, \theta, \alpha, \beta).$$

For fixed t, the inter-arrival times $\tau_{t1}, \ldots, \tau_{t,n+1}$, where $n = y_t$, are stochastically dependent, and the joint distribution factorizes as:

$$\begin{aligned} &p(\tau_{t1}, \ldots, \tau_{tn}, \tau_{t,n+1} | y_t = n, \theta, \alpha, \beta) \\ &= p(\tau_{t,n+1} | y_t = n, \theta, \alpha, \beta, \tau_{t1}, \ldots, \tau_{tn}) p(\tau_{t1}, \ldots, \tau_{tn} | y_t = n). \end{aligned}$$

The first n inter-arrival times are independent of all model parameters, and are determined only by the observed number of counts y_t. Due to well-known properties of a Poisson process, the first n arrival times are distributed as the order statistics of n Uniform $[0, 1]$-distributed random variables. Only the final inter-arrival time $\tau_{t,n+1}$ depends on the actual model parameters α, β and θ through the risk λ_t. Conditionally on y_t, only $n = y_t$ arrivals occur in $[0, 1]$, thus the $n + 1$-th arrival is known to occur after 1. Due to the memorylessness of the exponential distribution, the waiting time after having reached one has an exponential distribution with mean $1/\lambda_t$, and $\tau_{t,n+1} = 1 - \sum_{j=1}^{n} \tau_{tj} + \xi_t$, where $\xi_t \sim$ Exponential (λ_t). This leads to step (c1) to sample $\tau_{t1}, \ldots, \tau_{t,n+1}$.

To sample the indicators S from $p(S|\tau, y, \theta, \alpha, \beta)$, we use the fact that all indicators are conditionally independent given y, θ, α, β and τ:

$$p(S|\tau, y, \theta, \alpha, \beta) = \prod_{t=1}^{T} \prod_{j=1}^{y_t+1} p(r_{tj} | \tau_{tj}, \theta, \beta_t, \alpha).$$

Thus for each $t = 1, \ldots, T$, and each $j = 1, \ldots, y_t + 1$, the indicator r_{tj} is sampled independently from $p(r_{tj}|\tau_{tj}, \theta, \beta_t, \alpha)$. This density depends on the data only through τ_{tj} and depends on the model parameters θ, α and β_t only through the risk λ_t:

$$\text{pr}\{r_{tj} = k | \tau_{tj}, \theta, \beta_t, \alpha\} \propto p(\tau_{tj} | r_{tj} = k, \beta_t, \alpha, \theta) w_k,$$

where

$$p(\tau_{tj} | r_{tj} = k, \theta, \beta_t, \alpha) \propto \frac{1}{s_k} \exp\left(-\frac{1}{2} \left(\frac{-\log \tau_{tj} - \log \lambda_t - m_k}{s_k} \right)^2 \right).$$

The quantities $(w_k, m_k, s_k^2), k = 1, \ldots, R$ are the parameters of the finite mixture approximation (4), which are determined in advance and are held fixed throughout sampling. Step (c) for sampling the artificially missing sequences τ and S is easily implemented and involves draws from standard densities, only; namely sampling from uniform distributions, sampling from an exponential distribution and sampling from a discrete distribution.

Starting values for τ and S are obtained in the following way. Each component indicator r_{tj} is drawn uniformly from 1 to R. Steps (c1) to (c3) are used to sample starting values for $\tau_{t1}, \ldots, \tau_{tn}$ for each t, given the observed counts y_t. To obtain a starting value for $\tau_{t,n+1}$, we sample ξ_t from Exponential (λ_t) with $\lambda_t = y_t$, if $y_t > 0$. For all t, where $y_t = 0$, λ_t is set to a "small" value, in our examples we used $\lambda_t = 0.1$.

3.2 Adding a Rejection Step

Mixture auxiliary sampling could be seen as running a Metropolis-Hastings algorithm, based on proposing the unknown model parameters $\alpha, \beta_1, \ldots, \beta_T$ and θ from

Table 2: Expected acceptance rate (in percent) for a Metropolis-Hastings algorithm based on a mixture approximation with R components

λ	T	$(\lambda+1)T$	1	2	3	4	5	6	7	8	9	10
1	1	2	77.3	91.4	95.7	97.4	98.9	99.4	99.6	99.7	99.8	99.9
3	1	4	71.7	87.9	93.5	96.1	98.4	99.1	99.4	99.6	99.7	99.8
10	1	11	70.7	85.9	92.9	96.2	98.2	99.0	99.4	99.5	99.7	99.7
1	10	20	67.9	84.8	92.0	95.1	97.6	98.5	99.0	99.4	99.5	99.6
3	10	40	68.4	85.8	92.4	95.5	97.6	98.6	99.0	99.3	99.4	99.6
0.1	100	110	21.3	83.1	84.7	92.8	96.9	98.1	98.5	99.2	99.3	99.5
10	10	110	66.6	83.9	91.9	94.7	97.5	98.6	99.1	99.2	99.3	99.4
1	100	200	42.0	75.2	91.4	94.0	97.1	97.9	98.7	99.1	99.3	99.3
3	100	400	62.8	80.7	89.5	94.1	97.2	98.2	98.8	99.1	99.2	99.4
1	1000	2000	5.2	35.7	88.5	90.1	95.4	97.2	97.8	98.8	99.1	99.1

an approximate model, where in equation (3) the density $p_\varepsilon(\varepsilon_{tj})$ is substituted by the mixture approximation $q_{\varepsilon,R}(\varepsilon_{tj})$, however, without implementing the rejection step. A rejection step could be added, as suggested by several referees, however, developing another Metropolis Hastings algorithm is not our aim. First of all, the rejection step would deprive auxiliary mixture sampling of its simplicity. Second, introducing a mixture approximation to $p_\varepsilon(\varepsilon_{tj})$ is not necessary, when we implement a rejection step, because in this case we could use a single normal distribution as an approximation for $p_\varepsilon(\varepsilon_{tj})$, and avoid introducing the mixture indicators at all. The acceptance rate of this independence Metropolis sampler is quite high, as will be shown in the example below. Introducing a mixture approximation to $p_\varepsilon(\varepsilon_{tj})$ and a second level of data augmentation makes sense only, when this additional effort is compensated by avoiding the rejection step at all. Note that by increasing the number of components, the mixture approximation could be made arbitrarily accurate, as least theoretically, and we claim that such a rejection step is not essential, since the acceptance rate is close to 100 percent, if the number of components is large enough.

We evaluated this acceptance rate for a simple example, namely Bayesian inference for T independent observations $y_1, \ldots, y_T$ from the Poisson distribution Poisson (λ) under the prior $\lambda \sim$ Gamma (a_0, b_0), in which case the posterior of λ is known to arise from the Gamma $(a_0 + T\overline{y}, b_0 + T)$-distribution, which $\overline{y}$ being the sample mean. The augmented model, obtained after the first step of data augmentation reads:

$$-\log \tau_{tj} = \beta + \varepsilon_{tj}, \tag{6}$$

with $\beta = \log \lambda$. To evaluate, how the approximation error introduced in the second data augmentation step influences the acceptance rate, we consider a marginal two-step sampler without introducing the indicators, where we sample in a first step the inter-arrival times as in step (c1) and propose $\beta^{(new)}$ from the proposal density $q_R(\beta|\tau) \propto q_R(\tau|\beta)p(\beta)$, with $q_R(\tau|\beta)$ being the likelihood of an approximation to

model (6), obtained by substituting $p_\varepsilon(\varepsilon_{tj})$ by the mixture approximation $q_{\varepsilon,R}(\varepsilon_{tj})$:

$$q_R(\tau|\beta) = \prod_{t=1}^{T} \prod_{j=1}^{y_t+1} q_{\varepsilon,R}(-\log \tau_{tj} - \beta).$$

The acceptance rate depends on the ratio

$$r(\beta, \tau) = \frac{p(\tau|\beta)}{q_R(\tau|\beta)},$$

where $p(\tau|\beta)$ is the likelihood of the exact augmented model (6):

$$p(\tau|\beta) = \prod_{t=1}^{T} \prod_{j=1}^{y_t+1} p_\varepsilon(-\log \tau_{tj} - \beta).$$

The acceptance rate is random, depending both on the new draw $\beta^{(new)}$ as well as on draws $\beta^{(old)}$ and τ from the stationary distribution $p(\beta, \tau|y)$, which is known explicitly for this example. We determine the expected acceptance rate, where the expectation is taken with respect to the joint distribution of $\beta^{(new)}$, $\beta^{(old)}$ and τ:

$$\int \left[\int \min\left(1, \frac{r(\beta^{(new)}, \tau)}{r(\beta^{(old)}, \tau)}\right) q_R(\beta^{(new)}|\tau) d\beta^{(new)} \right] p(\tau|\beta^{(old)}, y) p(\beta^{(old)}|y) d\tau d\beta^{(old)}.$$

Table 2 reports this expected acceptance rate for various values of λ and T for increasing number of components for simulated data. First of all we find, that running a Metropolis-Hastings algorithm with $R = 1$, in which case $p_\varepsilon(\varepsilon_{tj})$ is approximated by a single Normal $(0.5772, 1.6625)$-distribution, and only the first augmentations step has to be implemented, is a reasonable alternative to auxiliary mixture sampling. By increasing the number of components, the acceptance rate evidently approaches 100%.

Note that the mixture approximation is applied to equation (6) not only once, but $\sum_{t=1}^{T} y_t + T$ times, thus on average $(\lambda+1)T$ approximations take place. Table 2 demonstrates, how the approximation error accumulates, when T as well as λ increase. For smaller number of components the acceptance rate rapidly decreases, as the number of expected approximations increases. For the 10 component mixture approximation, however, it remains above 99 percent, even for $\lambda = 1$ and $T = 1000$, where the expected number of approximations is equal to 2000.

3.3 Auxiliary Mixture Sampling for State Space Modelling of Time Series of Counts

To illustrate the practical application of auxiliary mixture sampling, we consider in detail state space modelling of time series of small counts as introduced by West et al. (1985) and Harvey and Fernandes (1989). In its most general form, the model reads:

$$\begin{aligned} &y_t|\alpha, \beta_t \sim \text{Poisson}\left(\exp(Z_t^1\alpha + Z_t^2\beta_t)\right), \\ &\beta_t = F\beta_{t-1} + c + w_t,\ w_t \sim \text{Normal}\left(0, Q\right), \end{aligned}$$

where β_t is a latent Markov process. The expectation $\mathrm{E}(\beta_t|\beta_{t-1}) = F\beta_{t-1} + c$ is linear in β_{t-1}, whereas the variance-covariance matrix is $\mathrm{Var}\,(\beta_t|\beta_{t-1}) = Q$. The matrices F, Q and the vector c may be known, or may depend on unknown model parameters θ. A simple example is the local level model which reads:

$$\begin{aligned} y_t|\mu_t &\sim \text{Poisson}\left(\exp(\mu_t)\right), \\ \mu_t &= \mu_{t-1} + w_t, \quad w_t \sim \text{Normal}\left(0, \theta\right), \end{aligned}$$

with θ being the only unknown model parameter.

Prior to the advent of Markov chain Monte Carlo methods, various approximation methods have been suggested in the literature to cope with the estimation problem for state space model for time series of counts. An approach that is related to, but different from MCMC methods is Monte Carlo EM estimation as implemented by Chan and Ledolter (1995). Another rather popular approximation method is based on assuming natural conjugate priors for β_t, based on discounting information from the past. Such methods have been studied in Harvey and Fernandes (1989) for state space models for time series of counts and qualitative observations, and in West et al. (1985) for the general dynamic linear model. Alternative approximate approaches which also allow for smoothing are based on the posterior mode filter of Fahrmeir (1992) and the integration-based Kalman-filter of Frühwirth-Schnatter (1994a). Each of these approximation methods is likely to introduce an approximation error of unknown magnitude, that is not reducible by increasing the computational effort of the investigator. A first attempt to compute the exact likelihood function for the Poisson local level model is reported in Kashiwagi and Yanagimoto (1992), which is basically an application of the numerical integration filter of Kitagawa (1987), and therefore limited to one- or two-dimensional state vectors. An advantage of Markov chain Monte Carlo methods in comparison to any of these methods, first of all lies in general in the fact that increasing the computational effort leads to increased accuracy of the algorithm. Second, the Markov chain Monte Carlo approach suggested in this paper allows for rather high-dimensional state vectors.

Application of the first data augmentations steps described above introduces a total of $n_t = y_t + 1$ inter-arrival times $\tau_{tj}, j = 1, \ldots, n_t$ for each of the T count observations $y_t, t = 1, \ldots, T$. The second data augmentation step introduces a component indicator r_{tj} for each of the $T + \sum_{t=1}^{T} y_t$ inter-arrival times τ_{tj}. After conditioning on all inter-arrival times as well as the component indicators, we end up with the following observation equation which is linear in the state vector β_t and has a normal observation error with known variance:

$$-\log \tau_{tj}|\mu_t, r_{tj} = Z_t^1\alpha + Z_t^2\beta_t + m_{r_{tj}} + \varepsilon_{tj}, \quad \varepsilon_{tj} \sim \text{Normal}\left(0, s^2_{r_{tj}}\right).$$

If we define a multivariate observation vector $\tilde{y}_t$ of dimension $n_t = y_t + 1$ as:

$$\tilde{y}_t = \begin{pmatrix} -\log \tau_{t1} - m_{r_{t1}} \\ \vdots \\ -\log \tau_{t,n_t} - m_{r_{t,n_t}} \end{pmatrix},$$

the augmented model may be written in the following linear Gaussian state space form:

$$\tilde{y}_t = \tilde{Z}_t^1 \alpha + \tilde{Z}_t^2 \beta_t + \varepsilon_t,\ \varepsilon_t \sim \text{Normal}\left(0, R_t\right), \tag{7}$$

$$\beta_t = F\beta_{t-1} + u_t + w_t,\ w_t \sim \text{Normal}\left(0, Q\right). \tag{8}$$

R_t is a diagonal matrix containing the variances of the mixture components, $R_t = \text{Diag}\left(s^2_{r_{t1}}, \ldots, s^2_{r_{t,n_t}}\right)$. $\tilde{Z}_t^1$ and $\tilde{Z}_t^2$ are matrices with n_t rows, containing n_t copies of the design matrices Z_t^1 and Z_t^2:

$$\tilde{Z}_t^1 = \begin{pmatrix} Z_t^1 \\ \vdots \\ Z_t^1 \end{pmatrix}, \qquad \tilde{Z}_t^2 = \begin{pmatrix} Z_t^2 \\ \vdots \\ Z_t^2 \end{pmatrix}.$$

Thus for a state space model for count data, application of the two data augmentation steps described above leads to a partially Gaussian state space model for repeated measurements, where the transition equation is the same as for the original Poisson state space model. The Poisson observation equation for the single count observation y_t, however, is substituted by a Gaussian observation equation with the multivariate observation vector $\tilde{y}_t$ appearing as repeated measurements.

The three-block auxiliary mixture sampler described in Subsection 3.1 works as follows:

(a) Multi-move sampling for the whole sequence $\alpha, \beta_0, \ldots, \beta_T$ by forward-filtering-backward sampling as in Frühwirth-Schnatter (1994b), Carter and Kohn (1994), de Jong and Shephard (1995), or Durbin and Koopman (2002) for the conditionally Gaussian state space form (7) and (8).

(b) Sample θ conditional on knowing α, β, τ and S from the conditionally Gaussian state space form (7) and (8).

(c) For each $t = 1, \ldots, T$, compute $\log \lambda_t = Z_t^1 \alpha + Z_t^2 \beta_t$, and sample the inter-arrival times $\{\tau_{tj}, j = 1, \ldots, y_t + 1\}$ and the component indicators $r_{tj}, j = 1, \ldots, y_t + 1$ as described in step (c).

The precise details in step (b) depend on the specific state space form. If Q is an unrestricted variance-covariance matrix, than Q is sampled from an inverted Wishart distribution. If only some diagonal elements of Q are unknown as for the basic structural model to be considered in Section 4, these parameters are sampled independently from inverted Gamma distributions.

3.4 Auxiliary Mixture Sampling for other Parameter-driven Models of Count Data

The auxiliary mixture sampler is useful also for other parameter-driven models of Poisson counts. The introduction of the two latent sequences τ and S eliminates non-normality and non-linearity, whenever the log intensity is linear in the unknown model parameters. To implement step (a) and (b) of auxiliary mixture sampling for

a particular model, we may exploit any result that is available for Markov chain Monte Carlo estimation of this particular model class within the Gaussian family. We only need to implement step (c), to sample the two latent sequences τ and S, where we condition on α, β and θ. To sample the inter-arrival times τ_{tj} we only need to know the observed counts and the conditional mean λ_t, whereas to sample the component indicator r_{tj} we need to know τ_{tj} and λ_t. Although λ_t depends on α, β and θ in a specific way described by the model, step (c) is independent of the specific structure of the model, once we determined λ_t.

Assume for further illustration, that we are fitting a random-effects model to panel count data y_{it}, $i = 1, \ldots, N, t = 1, \ldots, T$, as in Chib et al. (1998), who considered a model with multiple random effects based on the Poisson distribution $y_{it} \sim \text{Poisson}\left(\exp(\lambda_{it})\right)$. At each sweep of the auxiliary mixture sampler, each count observation y_{it} is augmented by inter-arrival times $\tau_{it,j}$ and indicators $r_{it,j}$ for $j = 1, \ldots, y_{it} + 1$. Through our data augmentation scheme, the random-effect model for count data reduces to the same random-effects model, however with $y_{it}+1$ repeated Gaussian measurements $-\log \tau_{it,j} - m_{r_{it,j}}, j = 1, \ldots, y_{it}+1$ with observation variance $s^2_{r_{it,j}}$. Implementation of step (a) and (b) are now standard. To sample τ and S in step (c), we only need to determine λ_{it}.

4 Application to Road Safety Data

We illustrate the usefulness of the proposed auxiliary mixture sampler on time series provided by the Austrian Road Safety Board. These time series are monthly counts of killed or injured pedestrians from 1987-2002 in Linz, which is the third largest town in Austria. We use series for two different age groups, children aged 6-10 and senior persons above 60. We are dealing with series of small counts not exceeding 5 respectively 15. A legal intervention intended to increase road safety took place during the observation period. More precisely, an amendment increasing priority for pedestrians became effective in Austria on October 1, 1994. Since then pedestrians who want to use a crosswalk have to be granted crossing. We are going to analyze separately for both age groups, the effect of this law on the risk λ_t in month t of being killed or seriously injured as a pedestrian living in Linz.

State space modelling as in Harvey and Durbin (1986) seems quite natural for these time series but the smallness of the counts makes an analysis using normal state space models clearly inappropriate. The number y_t of persons killed or seriously injured in time period t, follows the binomial distribution $\text{Bino}\left(e_t, \lambda_t\right)$, with the exposures e_t being equal the number of children or senior people living in Linz at time point t. As the risk λ_t is typically small, the binomial distribution is approximated by the Poisson distribution:

$$y_t \sim \text{Poisson}\left(e_t \lambda_t\right).$$

Both time series are modelled using a basic structural model for Poisson counts as in Durbin and Koopman (2001), where the risk λ_t is assumed to have a multiplicative trend as well as a multiplicative seasonal component:

$$\log(\lambda_t) = \mu_t + s_t. \tag{9}$$

In (9), it is assumed that μ_t is a stochastic trend, following a random walk with drift a_t:

$$\mu_t = \mu_{t-1} + a_{t-1} + w_{1t}, \qquad w_{1t} \sim \text{Normal}\left(0, \theta_1\right), \tag{10}$$

where $\mu_0 \sim \text{Normal}\left(y_1/e_1, 1\right)$. To capture the legal intervention effect, equation (10) is slightly modified, by including a level shift δ at the time point $t = t_{int}$, when legal amendments became effective:

$$\mu_t = \mu_{t-1} + a_{t-1} + \delta + w_{1t}. \tag{11}$$

In its most general form, the basic structural model assumes, that the drift a_t changes over time and follows a random walk itself:

$$a_t = a_{t-1} + w_{2t}, \qquad w_{2t} \sim \text{Normal}\left(0, \theta_2\right), \tag{12}$$

where $a_0 \sim \text{Normal}\left(0, 1\right)$. In the context of state space models, a_t is usually called the slope, as it determines the expected increase in the level of μ_{t+1} compared to μ_t.

Finally, $\exp(s_t)$ is a monthly multiplicative seasonal component generated by

$$s_t = -s_{t-1} - \cdots - s_{t-11} + w_{3t}, \qquad w_{3t} \sim \text{Normal}\left(0, \theta_3\right), \tag{13}$$

where $\sum_{j=0}^{11} s_{t-j} = 0$, and $(s_{-1}, \ldots, s_{-11})$ is an unknown initial pattern.

In equation (10) to (13), the parameters θ_1, θ_2 and θ_3 are unknown variances, which are either estimated from the data or assumed to be 0. Note that the stochastic trend reduces to a linear deterministic trend function with intercept μ_0 and slope a_0, if the variances θ_1 and θ_2 are zero. Choosing $\theta_3 = 0$ leads to a fixed seasonal pattern over the whole observation period, whereas choosing $\theta_3 > 0$ allows a smooth change in this pattern.

To estimate the model defined above, we rewrite it as a state space model, see for instance Harvey (1989) and Durbin and Koopman (2001). For a model with non-zero variances, for instance, the state vector β_t has 14 dimensions, namely $\beta_t = (\mu_t, a_t, s_t, \ldots, s_{t-11}, \delta)$, where only the first three components are actually dynamic. We now turn to the application of auxiliary mixture sampling scheme to estimate these models. Data augmentation through the mixture approximation leads to a partly dynamic model in the sense of Frühwirth-Schnatter (1994b) with $\theta = (\theta_1, \theta_2, \theta_3)$. In this model the variances can be sampled independently from inverse Gamma distributions, assuming inverted prior Gamma distributions on each variance. We choose $\theta_i \sim \text{InvGamma}\left(0.1, 0.001\right), i = 1, \ldots, 3$.

The Gibbs sampler described in Subsection 3.3 was run 12000 times with a burn in of 2000 runs. As the chain did not converge for the original formulation of the model we used a reparameterization where the seasonal component was non-centered as in Frühwirth-Schnatter (2004). The non-centered seasonal component $\tilde{s}_t$ is the standardized deviation of s_t from the initial seasonal pattern $\alpha = (s_{-1}, \ldots, s_{-11})$:

$$\tilde{s}_t = \frac{s_t - Z_t^1 \alpha}{\theta_4}, \tag{14}$$

$$\log(\lambda_t) = \mu_t + Z_t^1 \alpha + \theta_4 \tilde{s}_t, \tag{15}$$

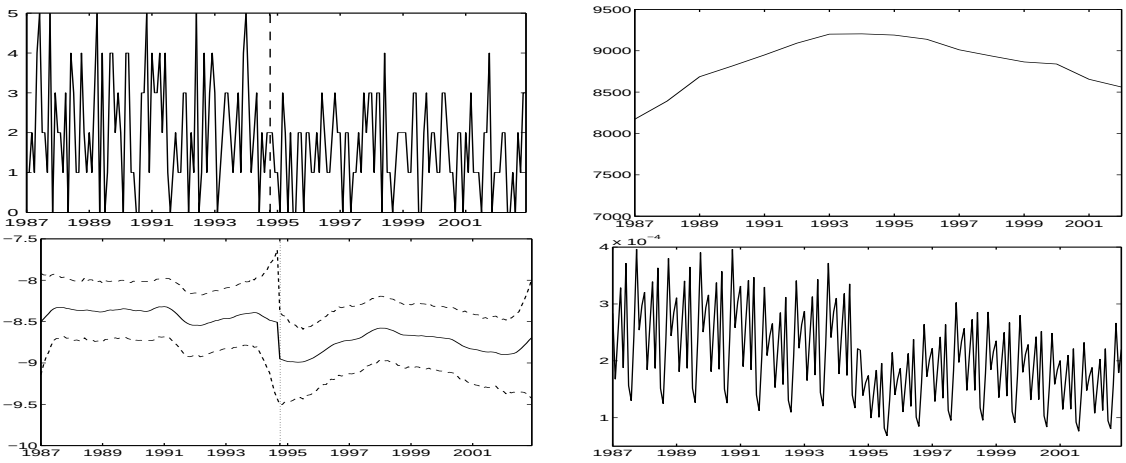

Figure 2: Above: Counts of killed or injured children (left hand side) and number of exposures (right hand side); below: estimated mean $\mu_t|y$ characterized by the posterior mean and 95% credible regions (left hand side) and estimated risk λ_t (right hand side)

where $\theta_4 = \sqrt{\theta_3}$. Z_t^1 is a row vector selecting the appropriate seasonal components for each time point t. For t being a multiple of 12, Z_t^1 is a row vector of -1s, otherwise all elements of Z_t^1 are 0, apart from the element in the column corresponding to the actual season, which takes the value 1. Introducing the state vector $\beta_t = (\mu_t, a_t, \tilde{s}_t, \ldots, \tilde{s}_{t-11}, \delta)$ and choosing $\theta = (\theta_1, \theta_2, \theta_4)$ led to a Gibbs sampler with quick convergence to the stationary distribution.

Figure 2 shows the observed counts, the number of exposures e_t, the smoothed level μ_t with point wise 95% credibility intervals and the risk λ_t for the children, Figure 3 shows the same for the senior series. The trend component a_t and the seasonal pattern s_t in the last year are shown in Figure 4 for both age groups.

The estimated risk is much larger for the children than for the senior people, and there is a pronounced decrease in risk for the children time series after the intervention. There are marked differences for the two series in the seasonal pattern: for the children series rates are significantly lower than the annual average in the holiday months July and August and higher June and October, for senior people there is solely a significant decrease in August.

Figure 5 shows the posterior density of θ_4 for both time series. Recall that θ_4 is defined as $\pm\sqrt{\theta_3}$, thus θ_4 has negative and positive sign with equal probability. Therefore, the posterior density of θ_4 is symmetric around 0. If the unknown variance θ_3 is systematically different from 0, then the posterior density of θ_4 is likely to be bimodal, otherwise, if θ_3 is close to 0, the posterior density of θ_4 will be centered around 0. The posterior density of θ_4 in Figure 5 shows for both time series that θ_4 is centered around zero, and we may conclude that for these time series the seasonal pattern is stable over time.

The drift term a_t is not significantly different from 0 in neither of the series, thus we are going to consider the simpler local level model with fixed seasonal pattern.

Tables 3 and 4 report point estimates as well as 95%-H.P.D. regions for the variance θ_1 and the intervention effect δ in the local level model with fixed seasonal

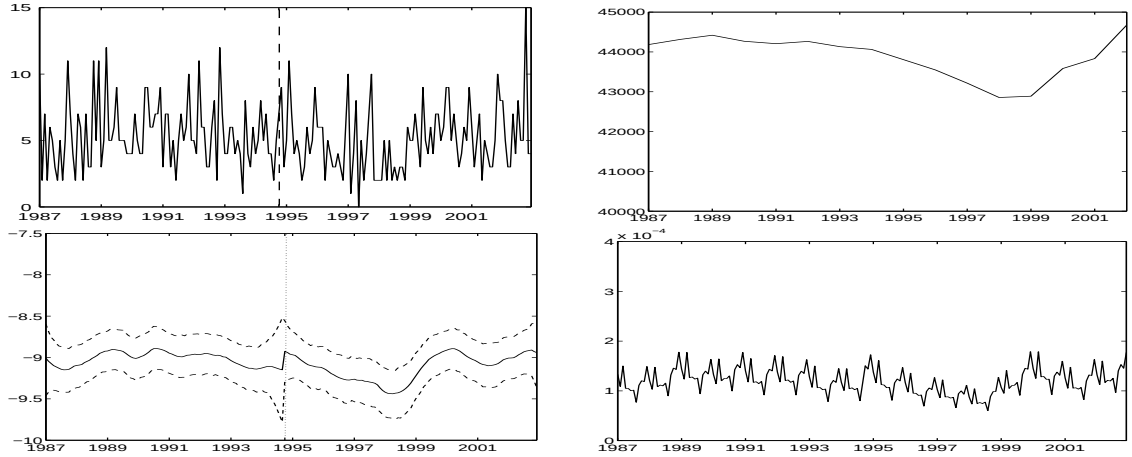

Figure 3: Above: Counts of killed or injured senior persons (left hand side) and number of exposures (right hand side); below: estimated mean $\mu_t|y$ with 95% credible regions (left hand side) and estimated risk $\lambda_t|y$ (right hand side)

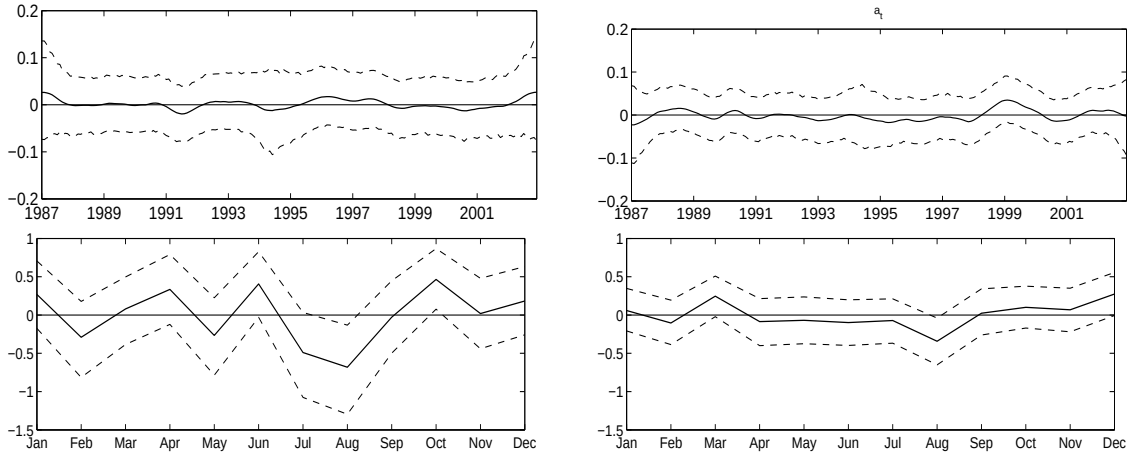

Figure 4: Posterior means of the drift a_t (above) and the seasonal component s_t in year 2002 (below) within 95% credible regions for the number of killed and injured children (left hand side) respectively senior people (right hand side)

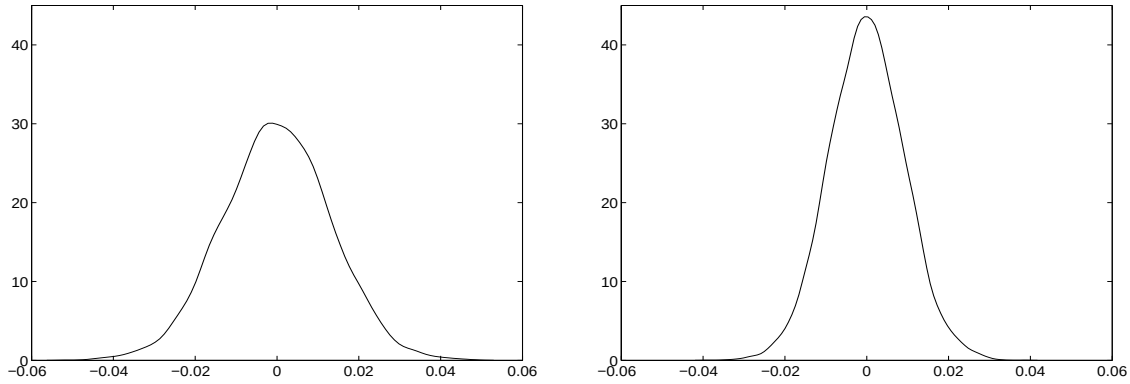

Figure 5: Kernel density estimate of the posterior distribution of θ_4 for children (left hand side) and senior people (right hand side)

Table 3: Parameter estimates for killed and injured children

Parameter	Mean	Std.dev	95%H.P.D. regions
θ_1	0.0022	0.0017	[0.0002, 0.0055]
δ	-0.4029	0.2453	[-0.9018, 0.0772]

pattern for both series. 95% HPD-regions are obtained as the shortest interval containing 95 percent of the simulations, see Chen et al. (2000) for more details. Whereas the process variances are nearly equal in both series, the intervention effect is negative for the children, but there is no intervention effect for the counts of killed or injured senior people.

5 Discussion and Concluding Remarks

The auxiliary mixture sampler suggested in this paper provides an important step toward operational Markov chain Monte Carlo estimation for a broad class of parameter-driven models of time series of counts. Some care, however, must be exercised with respect to parameterization issues, as straightforward Gibbs sampling often leads to convergence problems. Such problems are well-known for Gaussian random-effects model (Gelfand et al., 1995; van Dyk and Meng, 2001) and Gaussian state space models (Papaspiliopoulos et al., 2004; Frühwirth-Schnatter, 2004). For Poisson count data parameterization issues are also addressed in Chib et al. (1998). Our application demonstrates, that the mixing properties of auxiliary mixture sampling dramatically improves in cases, where the original parameterization leads to a slowly

Table 4: Parameter estimates for killed and injured senior people

Parameter	Mean	Std.dev	95%H.P.D. regions
θ_1	0.0020	0.0016	[0.0002, 0.0052]
δ	0.0417	0.1992	[-0.3401, 0.4512]

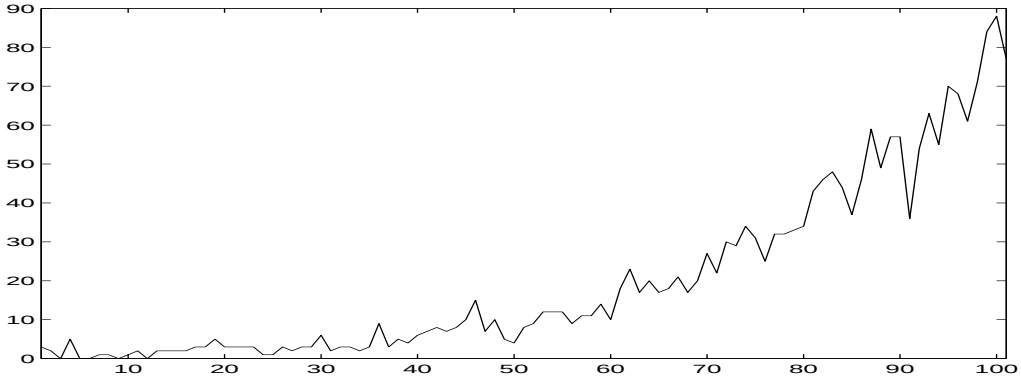

Figure 6: Simulated Poisson series with an exponential trend

mixing sampler, by using a non-centered parameterization similar to the one studied in Frühwirth-Schnatter (2004).

The auxiliary mixture sampler is mainly designed for, but not restricted to small counts, as larger counts can be safely treated with a normal approximation. It is, however, useful for time series including both small and large counts. To give an example, we generated a series y_t of length T=101 from a Poisson (λ_t)-distribution with exponential trend $\lambda_t = \exp(Z_t\alpha)$, and Z_t being evenly spaced from 0 to 5 and $\alpha = 0.9$. Figure 6 shows the simulated series. The Gibbs sampler was run without difficulties 12000 times with a burn-in of 2000, using the improper normal prior $p(\alpha) \propto 1$. It gave a posterior mean for α of 0.8967 with a standard error of 0.0058; the 95% credible interval was $[0.8852; 0.9073]$.

The auxiliary mixture sampler introduced in this paper is easily modified to deal with various extensions of the model structure, as any model with Poisson counts and some linear structure in the log intensity of the counts may be treated in the same way. If the latent process follows a t-distribution as in Chib and Winkelmann (2001), rather than a normal distribution, our estimation approach needs to be adapted only slightly along the lines of Shephard (1994), by expressing the t-distribution as a scale mixture of normals.

The results of the present paper are to some extent also useful outside the framework of data from a Poisson distribution. MacDonald and Zucchini (1997, p.68) note that a particularly useful model to capture overdispersion is a negative binomial distribution based on a hidden Markov process S_t, because such a model introduces overdispersion via the conditional distribution $p(y_t|S_t)$ as well as via the hidden Markov chain S_t. By writing the negative binomial distribution as an infinite mixture of Poisson distributions, an auxiliary mixture sampler is easily designed along the lines indicated in this paper for any model with a linear structure in the log of the mean.

Finally, as shown in Frühwirth-Schnatter and Frühwirth (2005), a similar auxiliary mixture sampler is feasible for parameter-driven models for other discrete-valued observations such as binary and multinomial data. Whereas the first data augmentation step for these type of observations is different and relies on the latent utility approach introduced by McFadden (1974), the second augmentation step interestingly involves a normal mixture approximation to the type I extreme

value distribution, which is equal to the density of minus log of an exponentially distributed variables, and is essentially the same as used in the present paper.

Acknowledgement

We want to thank Stefan Höglinger from the KfV, Austrian Road Safety Board for providing the data analyzed in Section 4.

References

Albert, J. (1992). A Bayesian analysis of a Poisson random-effects model. *American Statistician*, 46:246–253.

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

Chan, K. S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, 90:242–252.

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian computation*. Springer Series in Statistics. Springer-Verlag Inc, New York/Berlin/Heidelberg.

Chib, S., Greenberg, E., and Winkelmann, R. (1998). Posterior simulation and Bayes factors in panel count data models. *Journal of Econometrics*, 86:33–54.

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

Chib, S. and Winkelmann, R. (2001). Markov chain Monte Carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19:428–435.

Cox, D. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, 8:93–115.

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

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

Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (Disc: p29-56). *Journal of Royal Statistical Society, Series B*, 62:3–56.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *Journal of the American Statistical Association*, 87:501–509.

Frühwirth-Schnatter, S. (1994a). Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering. *Statistics and Computing*, 4:259–269.

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

Frühwirth-Schnatter, S. (2004). Efficient Bayesian parameter estimation for state space models based on reparameterizations. In Harvey, A., Koopman, S. J., and Shephard, N., editors, *State Space and Unobserved Component Models: Theory and Applications. Proceedings of a Conference in Honour of James Durbin*, pages 123–151, Cambridge. Cambridge University Press.

Frühwirth-Schnatter, S. and Frühwirth, R. (2005). Auxiliary mixture sampling with applications to logistic models. Research Report IFAS 2005-XX, http://www.ifas.jku.at/.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika*, 85:215–227.

Gelfand, A., Sahu, S. K., and Carlin, B. P. (1995). Efficient parametrisations for normal linear mixed models. *Biometrika*, 82:479–488.

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 Royal Statistical Society, Series A*, 149:187–227.

Harvey, A. C. and Fernandes, C. (1989). Time series models for count or qualitative observations (C/R: p418-422). *Journal of Business and Economic Statistics*, 7:407–417.

Kashiwagi, N. and Yanagimoto, T. (1992). Smoothing serial count data through a state-space model. *Biometrics*, 48:1187–1194.

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

Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series (C/R: p1041-1063; C/R: V83 p1231). *Journal of the American Statistical Association*, 82:1032–1041.

Lenk, P. J. and DeSarbo, W. S. (2000). Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika*, 65:93–119.

Leroux, B. G. and Puterman, M. L. (1992). Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics*, 48:545–558.

MacDonald, I. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman & Hall Ltd, London, monographs on statisticsand applied probability edition.

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

Papaspiliopoulos, O., Dellaportas, P., and Roberts, G. (2004). Bayesian inference for non-gaussian ornstein-uhlenbeck stochastic volatility processes. *Journal of Royal Statistical Society, Series B*, 66:To appear.

Robert, C. and Casella, G. (1999). *Monte Carlo statistical methods*. Springer Series in Statistics. Springer-Verlag Inc, New York/Berlin/Heidelberg.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika*, 81:115–131.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84:653–667.

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

van Dyk, D. and Meng, X.-L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10:1–50.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A., and Gelfand, A. E. (1994). Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. *Applied Statistics*, 43:201–221.

West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (C/R: p84-97). *Journal of the American Statistical Association*, 80:73–83.

Zeger, S. and Karim, M. (1991). Generalized linear models with random effects: a Gibbs sampling approach. *Journal of the American Statistical Association*, 86:79–86.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, 75:621–629.

Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, 44:1019–1031.