Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2004-01

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

Sylvia Frühwirth-Schnatter and Helga Wagner

August 2004

Abstract

In this article we consider parameter-driven models of time series of small counts, where the observed value follows an inhomogeneous Poisson process, with the mean changing over time according to a latent process. Estimation of these models is carried out within a Bayesian framework using data augmentation and MCMC methods. We suggest a new MCMC sampler, which possesses a Gibbs transition kernel, where we draw from full conditional distributions belonging to standard distribution families, only. Emphasis lies on application to state space modelling of small count. Nevertheless we show that our Gibbs sampling approach 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, Gibbs sampling, partially Gaussian state space models

1 Introduction

Applied statisticians commonly have to deal with time series of counts, recording 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 pairwise independent. To account for the dependency 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 classified into parameter-driven and observation-driven models. For observation driven models, the conditional distribution of y_t is specified as a function of the past observations $y_{t-1}, y_{t-2}, \ldots$, see for instance Zeger and Qaqish (1988). In this article we consider parameter-driven models, where the conditional distribution of y_t is allowed to change over time according to a latent process. This latent process could be a hidden Markov chain as in Leroux and Puterman (1992), or random effects as in Albert (1992). Smooth changes of the conditional distribution of y_t through state-space models have been considered by, among others, West et al. (1985), and Harvey and Fernandes (1989). Alternatively, a latent stationary autoregressive process has been introduced into the generalized linear model by Zeger (1988) and Chan and Ledolter (1995).

Estimation of parameter-driven Poisson time series models is known to be a challenging problem. In fact, estimation of these models using maximum likelihood estimation is hampered by the fact that the marginal likelihood, where the latent process is integrated out, is in general 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 MCMC 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 MCMC 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 turns make 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 that *straightforward Gibbs sampling* of all parameters, requiring only random draws from standard distributions such as multivariate normals, inverse Gamma, exponential and discrete distributions with a few categories is feasible for most of the parameter-driven models for time series of counts suggested in the literature so far. 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 a suitably chosen Poisson process. The introduction of this first sequence eliminates the non-linearity of the observation equation, whereas the non-normality of the error term, which follows a log Exponential (1)-distribution, remains. The log exponential distribution is 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 to the density of a $\log \chi^2$-distribution in the context of stochastic volatility models. By introducing the component indicator of this normal mixture 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 new and rather general Gibbs sampling scheme for parameter-driven models based on the Poisson distribution. 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

This section introduces a new method of data augmentation for parameter-driven models based on the Poisson distribution that will be exploited in the following section to implement straightforward Gibbs sampling for various time series models for count data. This data augmentation scheme is based on introducing two sequences of artificially missing data, leading to a Gaussian model, once we condition on the missing data. Thus we are able to show that any parameter-driven models based on the Poisson distribution may be regarded as a partially Gaussian model in the sense of Shephard (1994).

2.1 Model Specification

Let $y_1, \ldots, y_T$ be a sequence of count data. 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 $y = (y_1, \ldots, y_T)$:

$$p(\alpha, \beta_1, \ldots, \beta_T|\theta, y) \propto p(\alpha, \beta_1, \ldots, \beta_T|\theta) \prod_{t=1}^{T} p(y_t|\lambda_t). \tag{3}$$

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, in contrast to $p(\alpha, \beta_1, \ldots, \beta_T|\theta, y)$, is a joint 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. The first step of data augmentation introduces for each t, $t = 1, \ldots, T$, the inter-arrival times τ_{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:

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

This may be reformulated as following linear model:

$$\log \tau_{tj}|\alpha, \beta_t = -Z_t^1\alpha - Z_t^2\beta_t + \varepsilon_{tj}, \quad \varepsilon_{tj} \sim \log(\text{Exponential}\,(1)). \tag{4}$$

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 (4), which is non-normal, but where the mean of the observation equation is linear in the unknown model parameters $\alpha, \beta_1, \ldots, \beta_T$:

$$\text{E}(\log \tau_{tj}|\alpha, \beta_t) = -Z_t^1\alpha - Z_t^2\beta_t - 0.57722. \tag{5}$$

2.3 Step 2: Data augmentation through a Mixture Approximation

As discussed above, 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 (4) follows a log Exponential (1)-distribution which is independent of any unknown model parameter. To obtain a model that is conditionally Gaussian, we start by approximating the non-normal density of $\varepsilon_{tj} \sim \log(\text{Exponential}\,(1))$ by a normal mixture of R components with parameters m_r and s_r for the r-th component:

$$p(\varepsilon_{tj}) = \exp\{\varepsilon_{tj} - e^{\varepsilon_{tj}}\} \approx \sum_{r=1}^{R} w_r f_{\text{N}}(\varepsilon_{tj}; m_r, s_r^2). \tag{6}$$

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 χ^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 and are tabulated in Table 1 for

Table 1: Normal mixture approximation of the density of the log Exponential (1)-distribution (5 components)

r	1	2	3	4	5
w_r	0.2924	0.2599	0.2480	0.1525	0.0472
m_r	0.0982	-1.5320	-0.7433	0.8303	-3.1428
s_r^2	0.2401	1.1872	0.3782	0.1920	3.2375

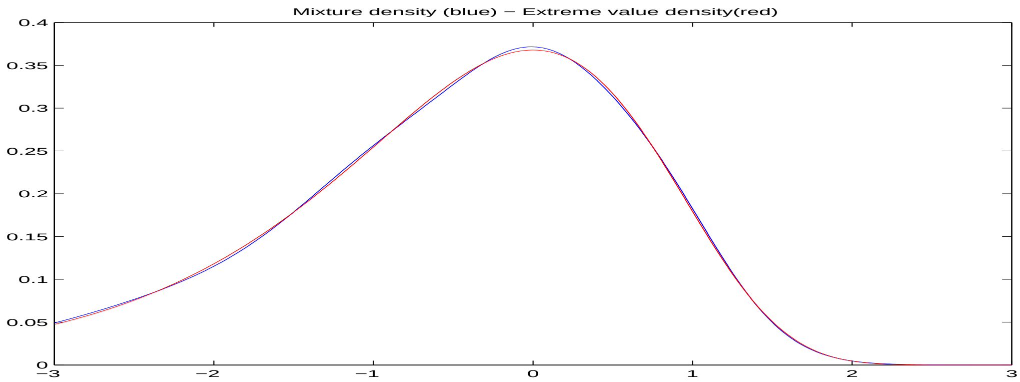

Figure 1: Comparing the density of the log(Exponential (1))-distribution with a normal mixture approximation with 5 components

$R = 5$, a number that we found to be sufficiently large in practice. For illustration, Figure 1 compares the true density with a normal mixture approximation based on 5 components.

Following Kim et al. (1998) and Chib et al. (2002), the mixture distribution (6) is regarded as the marginal distribution of a problem where additional to ε_{tj} the component indicators r_{tj} is observed. 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:

$$\log \tau_{tj}|\alpha, \beta_t, r_{tj} = -Z_t^1\alpha - Z_t^2\beta_t + m_{r_{tj}} + \varepsilon_{tj}, \quad \varepsilon_{tj} \sim \text{Normal}\left(0, s^2_{r_{tj}}\right). \quad (7)$$

Consequently, the conditional posterior $p(\alpha, \beta_1, \ldots, \beta_T|\theta, \tau, S, y)$, which is proportional to:

$$\begin{aligned} & p(\alpha, \beta_1, \ldots, \beta_T|\theta, \tau, S, y) & (8) \\ & \propto p(\alpha, \beta_1, \ldots, \beta_T|\theta) \prod_{t=1}^{T} \prod_{j=1}^{y_t+1} f_{\mathrm{N}}(\varepsilon_{tj}; \log \tau_{tj} - m_{r_{tj}} + Z_t^1\alpha + Z_t^2\beta_t, s^2_{r_{tj}}), \end{aligned}$$

is a multivariate normal density, which is easy to sample from.

3 Gibbs 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 has 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). A major difficulties with any of the existing MCMC 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 turns make it necessary to define suitable proposal densities in rather high-dimensional parameter spaces. Single-move sampling for this type of models is known to be potentially very inefficient, see e.g. Shephard and Pitt (1997).

In Section 2, we were able to show that any parameter-driven models based on the Poisson distribution may be regarded as a partially Gaussian model in the sense of Shephard (1994). This very useful result will be exploited in this section to implement straightforward Gibbs sampling for rather general parameter-driven models for time series of counts.

3.1 The Basic Four-block Gibbs Sampler

Select a starting value for the unknown model parameter θ, the component indicators $S = \{r_{tj}, j = 1, \dots, y_t + 1, t = 1, \dots, T\}$, and the inter-arrival times $\tau = \{\tau_{tj}, j = 1, \dots, y_t + 1, t = 1, \dots, T\}$ and repeat the following steps:

(a) Multi-move sampling of α and the *whole* sequence $\beta = \{\beta_1, \dots, \beta_T\}$ from the multivariate normal distribution (8), conditional on knowing τ, S, θ and y;

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

(c) sample the inter-arrival times $\tau = \{\tau_{tj}, j = 1, \dots, y_t + 1, t = 1, \dots, T\}$ conditional on knowing y, θ, α and β;

(d) sample the component indicators r_{tj} for each $\tau_{tj}, j = 1, \dots, y_t + 1, t = 1, \dots, T$.

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. Steps (c) and (d), however, deserve detailed investigation.

3.1.1 Sampling the inter-arrival times

The inter-arrival times $\{\tau_{tj}, j = 1, \dots, y_t + 1\}$ are independent for different time points t, given y, S, θ, α and β:

$$p(\tau|y, \theta, \alpha, \beta, S) = \prod_{t=1}^{T} p(\tau_{t1}, \dots, \tau_{t,y_t}, \tau_{t,y_t+1}|y_t, \theta, \alpha, \beta). \tag{9}$$

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

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

Note that the first n inter-arrival times are independent of all model parameters and in particular of the component indicator S, and are determined only by the observed number of counts y_t. 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, but is also independent of the component indicator S. 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. Therefore to sample $\tau_{t1}, \ldots, \tau_{tn}$ we perform the following steps:

(c1) sample n uniform random numbers $u_{t1}, \ldots, u_{tn}$;

(c2) sort them to obtain the first n arrival times $u_{t,(1)}, \ldots, u_{t,(n)}$;

(c3) define the inter-arrival times τ_{tj} as the increments of the arrival times:

$$\tau_{tj} = u_{t,(j)} - u_{t,(j-1)}, \quad j = 1, \ldots, n$$

where $u_{t,(0)} := 0$.

Conditionally on $y_t = n$ and $\tau_{t1}, \ldots, \tau_{tn}$, the last arrival time $\tau_{t,n+1}$ has an exponential distribution with mean $1/\lambda_t$, truncated at $1 - \sum_{j=1}^{n} \tau_{tj}$.

Therefore the final step reads:

(c4) define $\tau_{t,n+1} = 1 - \sum_{j=1}^{n} \tau_{tj} + \xi_t$, where $\xi_t \sim$ Exponential (λ_t).

If $y_t = 0$, we sample only a single arrival time τ_{t1} as $\tau_{t1} = 1 + \xi_t$, where $\xi_t \sim$ Exponential (λ_t).

3.1.2 Sampling the component indicators

The component indicators r_{tj} are mutually independent for different t as well as for different j, given y, τ, θ, α and β. For t, j fixed, the posterior of each component indicator r_{tj} depends on the data only through τ_{tj} and on the model parameters θ, α and β_t only through the risk λ_t:

$$\Pr\{r_{tj} = k | \tau_{tj}, \theta, \beta_t, \alpha\} \propto p(\tau_{tj} | r_{tj} = k, \beta_t, \alpha, \theta) w_k, \tag{10}$$

where

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

The quantities $(w_k, m_k, s_k^2), k = 1, \ldots, 5$ are the parameters of the finite mixture approximation tabulated in Table 1.

3.1.3 Starting values

Starting values for each component indicators r_{tj} are obtained as random draws from 1 to R. Steps (c1) to (c3) could be 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 use (c4) and sample ξ_t from Exponential (λ_t) with $\lambda_t = y_t$. For all t, where $y_t = 0$, λ_t can be set to a "small" value for λ_t, in our examples we used $\lambda_t = 0.1$.

3.1.4 Remarks

Sampling the inter-arrival times τ is carried out without conditioning on the component indicators S, whereas the component indicators are sampled conditional on knowing the inter-arrival times. Thus step (c) and (d) actually correspond to sampling (τ, S) jointly from the posterior $p(\tau, S|y, \theta, \alpha, \beta)$.

Both steps (c) and (d) for sampling the artificially missing sequences τ and S involve draws from standard densities, only; namely sampling from uniform distributions, sampling from an exponential distribution and sampling from a discrete distribution with $R = 5$ categories. Thus if Gibbs sampling is possible for a model, where the Poisson observation equation is substituted by a univariate normal distribution with mean $-\log \lambda_t$ and known variance, then the whole sampling scheme (a) to (d) is actually a Gibbs sampler for Poisson data.

3.2 Gibbs Sampling for State Space Models based on the Poisson Distribution

3.2.1 Introduction

For illustration of our Gibbs sampling scheme, we consider in detail state space modelling of time series of small counts. West et al. (1985) and Harvey and Fernandes (1989) extended the idea of generalized models based on the Poisson distribution to the framework of state space models which allows to introduce changing model parameters. In its most general form, the model reads:

$$y_t|\alpha, \beta_t \sim \text{Poisson}\left(\exp(Z_t^1 \alpha + Z_t^2 \beta_t)\right), \tag{11}$$

$$\beta_t = F\beta_{t-1} + u_t + w_t,\ w_t \sim \text{Normal}\,(0, Q)\,. \tag{12}$$

Prior to the advent of MCMC methods, various approximations 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 MCMC 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 MCMC approach suggested in this paper allows for rather high-dimensional state vectors.

3.2.2 Illustrative example: local level model for count data

A simple example of the model defined in equations (11) and (12) is the local level model for a univariate time series $\{y_1, \ldots, y_T\}$ of count data:

$$\begin{aligned} y_t|\mu_t &\sim \text{Poisson}\left(\exp(\mu_t)\right), \\ \mu_t &= \mu_{t-1} + w_t, \quad w_t \sim \text{Normal}\left(0, Q\right). \end{aligned} \tag{13}$$

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 μ_t and has a normal observation error with known variance:

$$\log \tau_{tj}|\mu_t, r_{tj} = -\mu_t + m_{r_{tj}} + \varepsilon_{tj}, \quad \varepsilon_{tj} \sim \text{Normal}\left(0, s^2_{r_{tj}}\right). \tag{14}$$

Thus for a state space model for Poisson count data, application of the two data augmentations steps described above leads to a partially Gaussian state space model for repeated measurements. The corresponding state space form reads

$$\begin{aligned} \tilde{y}_t|\mu_t &\sim \text{Normal}\left(-\tilde{Z}^2_t \mu_t, R_t\right) \\ \mu_t = \mu_{t-1} + w_t, &\quad w_t \sim \text{Normal}\left(0, Q\right) \end{aligned} \tag{15}$$

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 a multivariate observation vector $\tilde{y}_t$ given by:

$$\tilde{y}_t = \begin{pmatrix} \log \tau_{it,1} - m_{r_{it,1}} \\ \vdots \\ \log \tau_{it,n_t} - m_{r_{it,n_t}} \end{pmatrix}.$$

$\tilde{Z}^2_t$ is a column vector of ones of length $n_t = y_t + 1$ and R_t is a diagonal matrix containing the variances of the mixture components according to the sampled component indicators.

Gibbs sampling for this model is particularly simple. In step (a), multi-move sampling for $\mu = \{\mu_0, \ldots, \mu_T\}$ by forward-filtering-backward sampling involves only draws from one-dimensional normal distributions. For $t = T$,

$$\mu_T \sim \text{Normal}\left(\hat{\mu}_{T|T}, P_{T|T}\right), \tag{16}$$

where $\hat{\mu}_{T|T}$ and $P_{T|T}$ are the moments of the final filtering density $p(\mu_T|y, S, \tau)$. For $t < T$

$$\mu_t|\mu_{t+1} \sim \text{Normal}\left(\hat{\mu}_{t|T}, P_{t|T}\right) \quad t = T-1, \ldots, 0 \tag{17}$$

where $\hat{\mu}_{t|T}$ and $P_{t|T}$ are the moments of $p(\mu_t|\mu_{t+1}, y, S, \tau)$, see Frühwirth-Schnatter (1994b) for more details. Based on the conditionally conjugate prior $Q \sim \text{InvGamma}\,(c_0, C_0)$, the process variance Q is sampled in step (b) from an inverted Gamma distribution

$$Q|\mu \sim \text{InvGamma}\left(c_0 + \frac{T}{2}, C_0 + \frac{1}{2}\sum_{t=1}^{T}(\mu_t - \mu_{t-1})^2\right).$$

Step (c) and (d) are carried out as described above.

3.2.3 Gibbs Sampling Scheme for General State Space Models

For general state space models for count data, the unknown elements in the variance-covariance matrix Q in (12) in general constitute the unknown model parameter θ. The four-block Gibbs sampler works as follows. Select starting values for θ, and for the inter-arrival times and the component indicators as indicated in Subsection 3.1.3 and repeat the following steps:

(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) or de Jong and Shephard (1995) or by the sampler of Durbin and Koopman (2002) for the following conditionally Gaussian state space form:

$$\tilde{y}_t = -\tilde{Z}_t^1\alpha - \tilde{Z}_t^2\beta_t + \varepsilon_t,\ \varepsilon_t \sim \text{Normal}\,(0, R_t), \tag{18}$$

$$\beta_t = F\beta_{t-1} + u_t + w_t,\ w_t \sim \text{Normal}\,(0, Q). \tag{19}$$

The observation equation is Gaussian, with $\tilde{y}_t$ being the following multivariate observation vector of dimension $n_t = y_t + 1$:

$$\tilde{y}_t = \begin{pmatrix} \log\tau_{it,1} - m_{r_{it,1}} \\ \vdots \\ \log\tau_{it,n_t} - m_{r_{it,n_t}} \end{pmatrix}. \tag{20}$$

$\tilde{Z}_t^1$ 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:

$$\tilde{Z}_t^1 = \begin{pmatrix} Z_t^1 \\ \vdots \\ Z_t^1 \end{pmatrix}, \qquad \tilde{Z}_t^2 = \begin{pmatrix} Z_t^2 \\ \vdots \\ Z_t^2 \end{pmatrix}. \tag{21}$$

R_t is the diagonal matrix containing the variances of the mixture components, $R_t = \text{Diag}\left(s^2_{r_{it,1}}, \ldots, s^2_{r_{it,n_t}}\right)$.

(b) Sample θ conditional on knowing α, β, τ and S from the conditionally Gaussian state space form (18) and (19).

(c) For each $y_t, t = 1, \dots, T$, sample the inter-arrival times $\{\tau_{tj}, j = 1, \dots, y_t + 1\}$, conditional on knowing α, β and y as in Subsection 3.1.1. Sampling $\{\tau_{tj}, j = 1, \dots, y_t\}$ requires y_t draws from a uniform distribution, sampling τ_{t,y_t+1} requires a single draws from the Exponential (λ_t)-distribution with

$$\log \lambda_t = Z_t^1 \alpha + Z_t^2 \beta_t.$$

(d) Sample the component indicators r_{tj} for each $\tau_{tj}, j = 1, \dots, y_t + 1, t = 1, \dots, T$ as in Subsection 3.1.2 from the discrete distribution (10) with $\log \lambda_t$ being the same as in step (c).

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

3.3 Gibbs Sampling for other Parameter-driven Models of Count Data

The application of the four-block Gibbs sampler to state space modelling, discussed in the previous section, demonstrates two important features of our sampling scheme, which render it useful also for other parameter-driven models of counts.

First, the introduction of the two latent sequences τ and S eliminates non-normality and non-linearity caused by choosing the Poisson distribution as distributional law for the observations. Thus to extend a particular model class to count data, we may exploit any result that is available for MCMC estimation of this particular model class *within the Gaussian family*, when implementing step (a) and (b) of the four-block Gibbs sampler.

Second, in the sampling steps for the two latent sequences τ and S in step (c) and (d), where we condition on α, β and θ, knowledge of the conditional mean λ_t of the Poisson distribution of y_t is sufficient. Although λ_t depends on α, β and θ in a specific way described by the model, step (c) and (d) are independent of the specific structure of the model, once we determined λ_t. 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.

Assume, for further illustration, that we are fitting a random-effects model to panel count data y_{it}, $i = 1, \dots, N, t = 1, \dots, T$:

$$y_{it} \sim \text{Poisson}\left(\exp(\lambda_{it})\right), \tag{22}$$

as in Chib et al. (1998), who considered panel data models with multiple random effects. With our data augmentation scheme any random-effect model for small counts reduces to the same random-effects model for repeated Gaussian data. At each sweep of our Gibbs sampler, each count observation y_{it} is substituted by $y_{it} + 1$

repeated Gaussian measurements $\log \tau_{it,j} - m_{r_{it,j}}, j = 1, \ldots, y_{it} + 1$ with observation variance $s^2_{r_{it,j}}$. It is important to realize the following. First, these Gaussian measurements are completely determined by the most recent draws of the two latent sequences τ and S. Second, the unknown model parameters are conditionally independent of the original counts y_{it}, conditional on knowing the repeated Gaussian measurements $\log \tau_{it,j} - m_{r_{it,j}}, j = 1, \ldots, y_{it} + 1$. Third, once we conditioned on τ and S, the mean of each of the Gaussian observations $\log \tau_{it,j} - m_{r_{it,j}}$ is equal to $-\lambda_{it}$, with the model for λ_{it} being identical with the original model.

4 Applications

We illustrate the usefulness of the proposed Gibbs sampler on various data sets provided by the Austrian Road Safety Board . These data are monthly observations of numbers of fatal accidents of female drivers in our first example and numbers of killed or injured pedestrians of two age-categories in the second. We deal with series of small counts not exceeding 4 in the first, and respectively 5 and 15 in the second example. State space modelling seems quite natural for these data but the smallness of the counts makes an analysis using normal state space models as in Harvey and Durbin (1986) clearly inappropriate.

All series are modelled using a basic structural model for count data, i.e.

$$\log(\lambda_t) = \mu_t + s_t$$

where the level μ_t and slope a_t follow a random walk

$$\mu_t = \mu_{t-1} + a_{t-1} + w_{1t}, \qquad w_{1t} \sim \text{Normal}\,(0, \theta_1) \tag{23}$$

$$a_t = a_{t-1} + w_{2t}, \qquad w_{2t} \sim \text{Normal}\,(0, \theta_2)\,, \tag{24}$$

and the monthly seasonal component is generated by

$$s_t = -s_{t-1} - \cdots - s_{t-11} + w_{3t}, \qquad w_{3t} \sim \text{Normal}\left(0, \theta_3^2\right)$$

In both examples a legal intervention intended to increase road safety took place during the observation period. The intervention effect δ is modelled as a level shift at the time point $t = t_{int}$ when legal amendments became effective, so

$$\mu_t = \mu_{t-1} + a_{t-1} + \delta + w_{1t} \qquad \text{for } t = t_{int}.$$

The state vector β_t therefore has 14 dimensions, namely $\beta_t = (\mu_t, a_t, s_t, \ldots, s_{t-11}, \delta)$, where only the first three components are dynamic. Data augmentation through the mixture approximation leads to a partly dynamic model in the sense of Frühwirth-Schnatter (1994b) with process variances $\theta = (\theta_1, \theta_2, \theta_3^2)$ for the three dynamic components. In this model the process variances can be sampled independently from inverse Gamma distributions.

4.1 Example 1: Female Drivers in Rohrbach

In this example we analyze monthly counts of killed or seriously injured female drivers in the district of Rohrbach, which is a rural district typical for northern Austria. The time period covered was 1981 to 2003.

Table 2: Estimates of the parameters for killed or seriously injured female drivers

Parameter	Mean	Std.dev	95%H.P.D. regions
θ_1	0.0049	0.0058	[0.0001, 0.0168]
θ_2	0.0003	0.0002	[0.0001, 0.0007]
θ_3	0.0011	0.0216	[-0.0407, 0.0447]
δ	-0.1156	0.9055	[-1.8994, 1.6673]

This series allows to investigate the effect of Austrian seat belt legislation analogous to the analysis in Harvey and Durbin (1986). Use of the seat belt for front seat occupants was made compulsory in Austria in 1976, but violations were not prosecuted till July 1, 1984.

The Gibbs sampler described in Subsection 3.2.3 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 reparametrization where the seasonal component was noncentred for location as well as scale:

$$\log(\lambda_t) = \mu_t + Z_t^1 \alpha + \theta_3 \tilde{s}_t \tag{25}$$

Here the initial seasonal pattern is introduced as fixed effect, $\alpha = (s_{-1}, \dots, s_{-11})$, and Z_t^1 is a row vector selecting the appropriate initial value according to the season of time point t. For t being a multiple of 12, Z_t^1 is a row vector of -1, 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. The non-centered seasonal component $\tilde{s}_t$ is the standardized deviation of s_t from α:

$$\tilde{s}_t = \frac{s_t - Z_t^1 \alpha}{\theta_3}. \tag{26}$$

Introducing the non-centered state vector $\beta_t = (\mu_t, a_t, \tilde{s}_t, \dots, \tilde{s}_{t-11}, \delta)$ and choosing $\theta = (\theta_1, \theta_2, \theta_3)$ led to a Gibbs sampler with quick convergence to the stationary distribution.

Table 2 summarizes the obtained estimates for the model. Process variances for all components are low, particularly variances for the linear trend and the seasonal component are close to 0. Recall that θ_3 is the process standard error of the seasonal component. Due to noncentring estimates of θ_3 may also have negative signs. The variance of the seasonal component is not significantly different from 0, indicating a stable seasonal pattern over the observation period. The intervention effect is negative, though not significant. Thus prosecution of non-wearers of the seatbelt does not have a significant effect on female drivers in this rural area.

Figure 2 shows the observed counts with the exponentiated estimated level μ_t and pointwise 95% credible intervals, figure 3 the (multiplicative) trend $\exp(a_t)$ and a typical seasonal pattern $\exp(s_t)$. Obviously neither trend nor seasonal effects are significantly different from 1.

4.2 Example 2: Killed or injured pedestrians

In this application we study series of monthly counts of deaths or injured pedestrians from 1987-2002 in Linz, which is the third largest town in Austria. We use

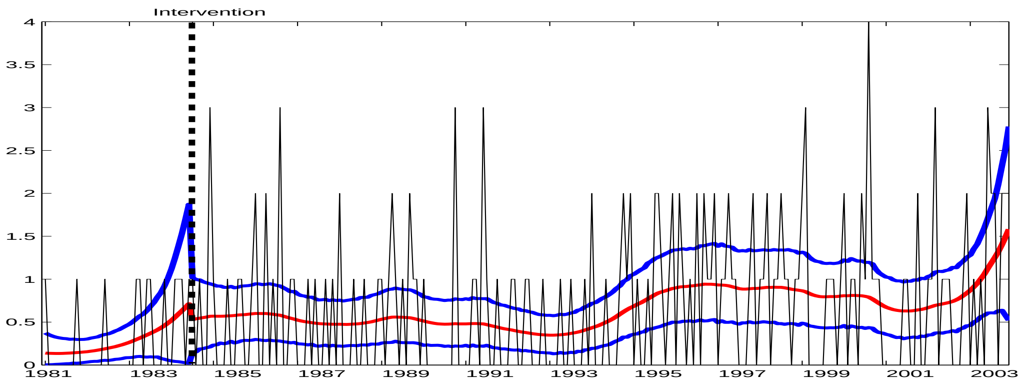

Figure 2: Counts of killed or seriously injured female drivers in Rohrbach with estimated rate (posterior means) within 95% credible regions

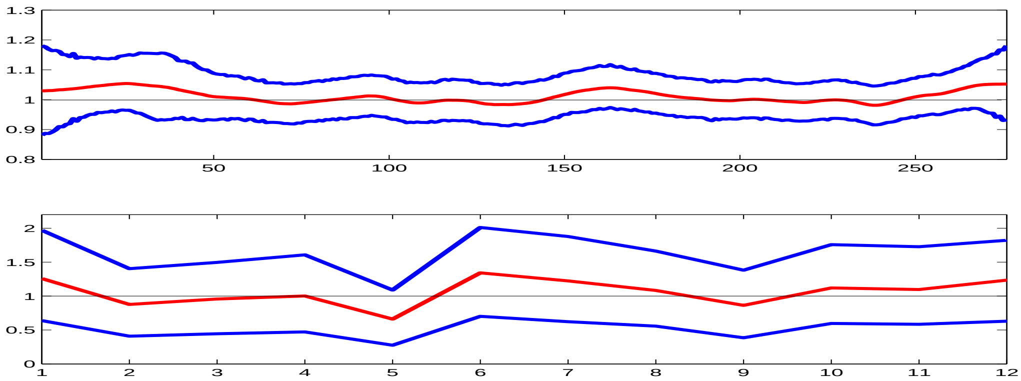

Figure 3: Trend and seasonal component (posterior means) within 95% credible regions

Table 3: Parameter estimates for killed and injured children and seniors

	Children			Senior people		
Parameter	Mean	Std.dev	95%H.P.D. regions	Mean	Std.dev	95%H.P.D. regions
θ_1	0.0042	0.0046	[0.0002, 0.0140]	0.0040	0.0043	[0.0002, 0.0121]
θ_2	0.0004	0.0002	[0.0001, 0.0007]	0.0003	0.0002	[0.0001, 0.0006]
θ_3	0.0002	0.0195	[-0.0378, 0.0387]	-0.0004	0.0154	[-0.0306, 0.0288]
δ	-0.5128	0.5779	[-1.6920, 0.5812]	0.2966	0.3700	[-0.4202, 1.0259]

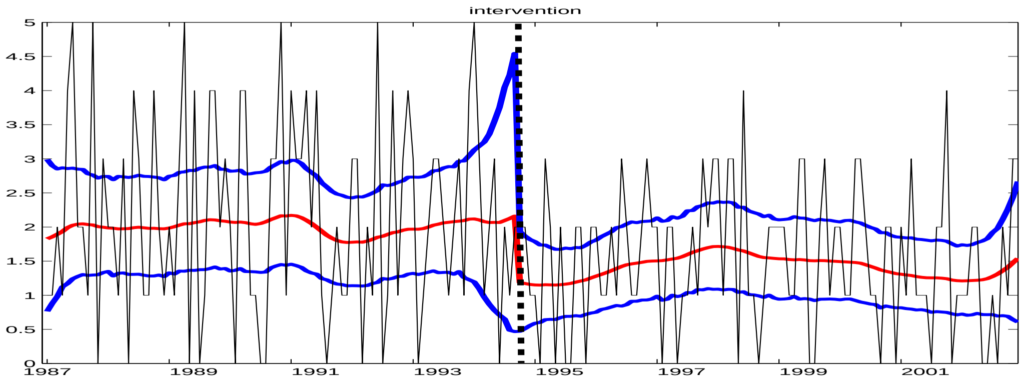

Figure 4: Counts of killed or injured children with estimated rate (posterior means) within 95% credible regions

series for two different age groups, children aged 6-10 and senior persons above 60. Legal intervention in this application concerns an amendment increasing priority for pedestrians which became effective on October 1, 1994: since then pedestrians who want to use a crosswalk have to be allowed a riskless crossing.

For both series we fitted a basic structural model including an immediate intervention effect. Again we had to use the reparametrized version of the sampler due to non convergence of the original version.

Figures 4 and 6 show the observed counts with the smoothed level and point wise 95% credibility intervals for both series. Table 3 reports point estimates as well as 95%-H.P.D. regions for all modell parameters. All components show an almost equal variability. As θ_3 is not significantly different from zero, also in these series the seasonal pattern is stable along the observation period.

The (multiplicative) trend component and typical seasonal patterns are shown in figures 5 and 7. The trend component is not significantly different from 1 in neither of the series. There are however marked differences for the two series in the seasonal patterns: 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.

As the main feature of the seasonal pattern in the children series is the decrease in holiday months July and August we fitted a simpler model with a dummy variable indicating summer holidays and omitting the insignificant trend component a_t. In this model the estimated variance of the level component was $\hat{\theta}_1 = 0.0018$,with an

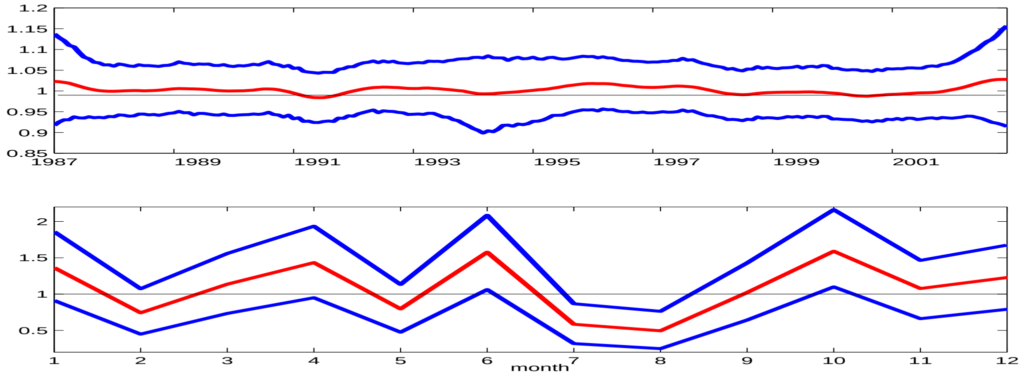

Figure 5: Killed or injured children: Trend and typical seasonal component (posterior means) within 95% credible regions

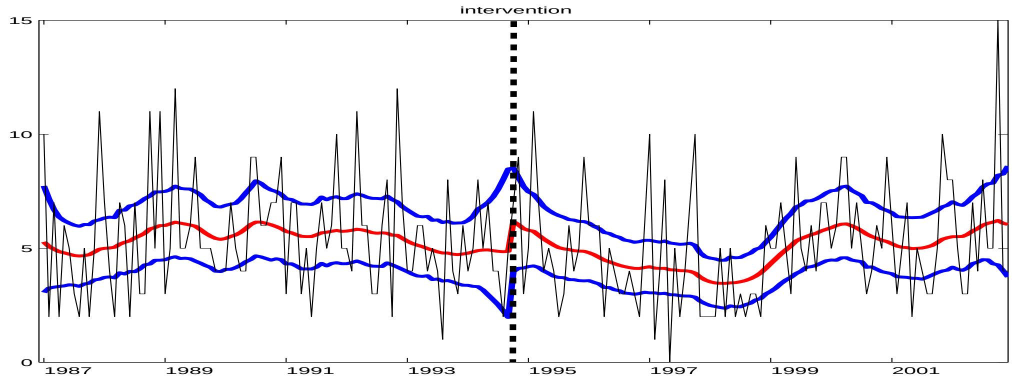

Figure 6: Counts of killed or injured senior persons with estimated rate (posterior means) within 95% credible regions

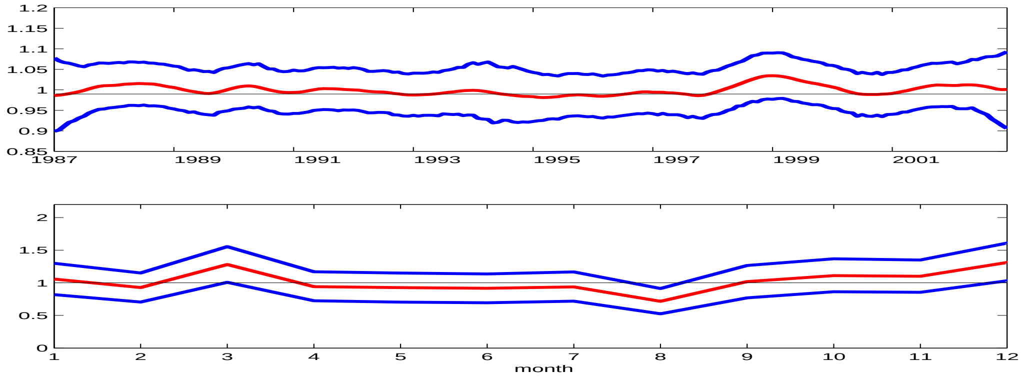

Figure 7: Senior persons: Trend and typical seasonal component (posterior means) within 95% credible regions

estimated intervention effect of $\hat{\delta} = -0.4006$. For the senior series a similar model with a dummy variable for August gave estimates $\hat{\theta}_1 = 0.0030$ and $\hat{\delta} = 0.1334$.

5 Concluding Remarks

The Gibbs sampler suggested in this paper provides an important step toward operational MCMC estimation for a broad class of parameter-driven models of time series of Poisson counts, as the sampler typically requires only draws from standard densities and no tuning of proposal densities is required.

Some care 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). As our application demonstrated, the mixing properties of our new Gibbs sampling scheme dramatically improves in cases, where the original parameterization leads to a slowly mixing sampler, by using a non-centered parameterization similar to the one studied in Frühwirth-Schnatter (2004).

The Gibbs sampler introduced in this paper is easily modified to deal with various extensions of the model structure. 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. Furthermore, the observations y_t may be regarded as realizations from a negative binomial distribution, which is an important alternative to the Poisson distribution that is able to capture overdispersion often present in count data. By writing the negative binomial distribution as an infinite mixture of Poisson distributions, an MCMC scheme is easily designed along the lines indicated in this paper.

Although we focused on count data throughout the paper, the main ideas are likely to be useful for constructing straightforward Gibbs sampling schemes for parameter-driven models for other discrete observations such as binary or multinomial data. This issue, however, will be pursued in more detail in a subsequent paper.

Acknowledgement

We are grateful to Rudi Frühwirth for providing the normal mixture approximation to the log exponential distribution which is in fact an essential ingredient for our Gibbs sampling scheme. Furthermore we want to thank Stefan H¨glinger 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.

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.

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. 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.

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.

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.