I²SDS - Decision Sciences
THE GEORGE WASHINGTON UNIVERSITY
SCHOOL OF BUSINESS

The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2014-2
February 8, 2014

Bayesian Modelling of Time Series of Counts

Refik Soyer
Department of Decision Sciences
The George Washington University, USA

Tevfik Aktekin
Department of Decision Sciences
University of New Hampshire, USA

Bumsoo Kim
Business Administration, Sogang Business School
Sogang University, Korea

Chapter 1

Bayesian Modelling of Time Series of Counts

Refik Soyer, Tevfik Aktekin and Bumsoo Kim

Abstract

In this chapter we consider time-series of correlated counts which often arise in finance, operations and marketing applications. We present a class of parameter-driven time-series models for counts using a Bayesian state-space approach. In so doing, we consider Poisson counts and present multivariate extensions to negative-binomial time-series. We develop Bayesian inference of these models using Markov chain Monte Carlo methods and present applications such as modelling of mortgage defaults, call center arrivals, and shopping trips.

1.1 Introduction

In time series analysis, observations under study can often be the number of occurrences of an event of interest in a given time interval, also referred to as count data. Such type of data can arise in numerous fields such as engineering, business, economics or epidemiology. For instance, observations under study can be the number of arrivals to a bank in a given hour, number of shopping trips of households in a week, number of mortgages defaulted from a particular pool in a given month, number of accidents in a given time interval or the number of deaths from a specific disease in a given year.

Studies in time series with focus on count data is scarce compared to those with continuous data. Many of such studies in the time series analysis literature consider a Poisson model for count data. A regression model where the observations are time series of counts has been introduced by Zeger (1988) where a quasi-likelihood method is used in order to estimate model parameters. Harvey and Fernandes (1989) assume a Gamma process on the stochastic evolution of the latent Poisson mean and propose extensions to other count data models such as the binomial, the multinomial and negative binomial distributions. Davis et al. (2000) develop a method in order to diagnose the latent factor embedded in the mean of a Poisson regression model and present asymptotic properties of model estimators. Davis et al. (2003) discuss the maximum likelihood estimation of a general class of observation-driven models of count data, also referred to as generalized autoregressive moving average models for counts and develop relevant theoretical properties. Freeland and McCabe (2004) introduce new methods of assessing the fit of the Poisson autoregressive model via the information matrix equality, provide properties of the maximum likelihood estimators and discuss further implications in residual analysis.

The Bayesian point of view has also been considered in the time series analysis of count data. Chib et al. (1998) introduce Markov chain Monte Carlo (MCMC) methods to estimate Poisson panel data models with multiple random effects and discuss implications of model fit for different Bayes factor estimators. Chib and Winkelmann (2001) propose a model which can take into account correlated count data via latent effects and show how the estimation method is practical even with high dimensional count data. Bayesian State space modeling of Poisson count data is considered by Fruehwirth-Schnatter and Wagner (2006) where MCMC via data augmentation techniques to estimate model parameters is used. Furthermore, Durbin and Koopman (2000) discuss the state space analysis of non-Gaussian time series models from both classical and Bayesian perspectives, apply an importance sampling technique for estimation and illustrate an example with count data. More recently, Santos et al. (2013) consider a non-Gaussian family of state space models with exact marginal likelihood with the Poisson is one of the special cases.

In this paper, we introduce a general class of Poisson time series models from a Bayesian state space modeling perspective where we propose different strategies for the stochastic Poisson rate which evolves over time. Such models are also referred to as parameter-driven time series models as described by Davis et al. (2003). As pointed out by Durbin and Koopman (2000), the attractive feature of the state space approach is that it allows the modeling of numerous sub-components to form an overall system of time series data. We also present an extension of our framework to multivariate counts where dependence between the individual counts is motivated by a common environment. This approach provides multivariate negative-binomial models for the time-series. In the proposed framework we develop Bayesian inference by using MCMC methods. In doing so, for each proposed model we discuss appropriate MCMC estimation techniques such as the Gibbs sampler, the Metropolis-Hastings algorithm and the forward filtering backward sampling algorithm. For a good introduction to most common MCMC practice see Smith and Gelman (1992) and Chib and Greenberg (1995), and for forward filtering backward sampling see Fruhwirth-Schnatter (1994).

A quick summary of our paper is as follows: First, we introduce a Bayesian state space model for count time series data, show how the parameters can sequentially be updated, discuss smoothing, filtering and forecasting and introduce further properties of the proposed model. Next, we discuss how covariate information can be incorporated in the Poisson rate and introduce MCMC methods to estimate model parameters. This is followed by a section dedicated to multivariate extensions of the basic model. The proposed approach and its extensions are illustrated by three examples from finance, operations and marketing. The last section concludes with some remarks.

1.2 A Discrete Time Poisson Model

In this section, we introduce a Poisson time series model whose stochastic rate evolves over time according to a discrete time Markov process. We refer to this model as the basis model whose extensions will be discussed in the sequel. We note that the our model is based on the framework proposed by Smith and Miller (1986) where the measurement model was exponential. Let N_t be the number of occurrences of an event in a given time interval t and let θ_t be the corresponding latent Poisson rate during the same time period. Given the rate θ_t, we assume that the number of occurrences of an event during period t is described by a discrete time non-homogeneous Poisson process with probability distribution of the following form

$$p(N_t|\theta_t) = \frac{\theta_t^{N_t} e^{-\theta_t}}{N_t!}. \tag{1.2.1}$$

It is assumed that N_ts are conditionally independent given θ_ts. Namely, the independent increments property holds only conditional on θ_t, but unconditionally, N_ts are correlated. (1.2.1) is the measurement equation of a discrete time Poisson state space model. For the time evolution of the latent rate process, θ_ts, we assume the following Markovian structure

$$\theta_t = \frac{\theta_{t-1}}{\gamma}\epsilon_t, \tag{1.2.2}$$

where $(\epsilon_t|D^{t-1}) \sim Beta[\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1}]$ with $\alpha_{t-1} > 0$, $0 < \gamma < 1$, and $D^{t-1} = \{N_1, \cdots, N_{t-1}\}$. In (1.2.2), γ acts like a discounting term and its logarithm can be considered to be the first order autoregressive component for the latent rates, θ_ts. It follows from (1.2.2) that there is an implied stochastic ordering between two consecutive rates, $\theta_t < \frac{\theta_{t-1}}{\gamma}$. It is also straight forward to show that the conditional distributions of consecutive rates are all scaled Beta densities, $(\theta_t|\theta_{t-1}, D^{t-1}) \sim Beta[\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1}; (0, \frac{\theta_{t-1}}{\gamma})]$, and are given by

$$p(\theta_t|\theta_{t-1}, D^{t-1}) = \frac{\Gamma(\alpha_{t-1})}{\Gamma(\gamma\alpha_{t-1})\Gamma(\{1-\gamma\}\alpha_{t-1})}\Big(\frac{\gamma}{\theta_{t-1}}\Big)^{\alpha_{t-1}-1}\theta_t^{\gamma\alpha_{t-1}-1}\Big(\frac{\theta_{t-1}}{\gamma} - \theta_t\Big)^{(1-\gamma)\alpha_{t-1}-1}. \tag{1.2.3}$$

The state equation (1.2.2) also implies that $E(\theta_t|\theta_{t-1}, D^{t-1}) = \theta_{t-1}$, in other words a random walk type of evolution in the expectation of the Poisson rates.

As a consequence of the measurement and state equations, it is possible to develop an analytically tractable sequential updating of the model if we assume that at time 0, $(\theta_0|D^0)$ is a gamma distribution as

$$(\theta_0|D^0) \sim Gamma(\alpha_0, \beta_0). \tag{1.2.4}$$

Given the inductive hypothesis

$$(\theta_{t-1}|D^{t-1}) \sim Gamma(\alpha_{t-1}, \beta_{t-1}), \tag{1.2.5}$$

a recursive updating scheme can be developed as follows. Using (1.2.3) and (1.2.5), we can obtain the distribution of θ_t given D^{t-1} as

$$(\theta_t|D^{t-1}) \sim Gamma(\gamma\alpha_{t-1}, \gamma\beta_{t-1}). \tag{1.2.6}$$

It follows from the above that $E(\theta_t|D^{t-1}) = E(\theta_{t-1}|D^{t-1})$, whereas $V(\theta_t|D^{t-1}) = \frac{V(\theta_{t-1}|D^{t-1})}{\gamma}$. In other words, as we move forward in time our uncertainty about the rate increases as a function of γ. Given the prior (1.2.5) and the Poisson observation model (1.2.1) we can obtain the filtering distribution of $(\theta_t|D^t)$ using the Bayes' Rule as

$$p(\theta_t|D^t) \propto p(N_t|\theta_t)p(\theta_t|D^{t-1}). \tag{1.2.7}$$

The above implies that

$$p(\theta_t|D^t) \propto \theta_t^{\gamma\alpha_{t-1}+N_t-1} e^{-(\gamma\beta_{t-1}+1)\theta_t},$$

that is, the filtering distribution of the Poisson rate at time t is a gamma density

$$(\theta_t|D^t) \sim Gamma(\alpha_t, \beta_t), \tag{1.2.8}$$

where the recursive updating of model parameters is given by $\alpha_t = \gamma\alpha_{t-1} + N_t$ and $\beta_t = \gamma\beta_{t-1} + 1$. The fact that $(\theta_t|D^t)$ is available in closed form is an attractive feature of the proposed model from a practical point of view. In addition, one-step ahead forecasting distribution of counts at time t given D^{t-1} can be obtained via

$$p(N_t|D^{t-1}) = \int_0^\infty p(N_t|\theta_t)p(\theta_t|D^{t-1})d\theta_t, \tag{1.2.9}$$

where $(N_t|\theta_t) \sim Poisson(\theta_t)$ and $(\theta_t|D^{t-1}) \sim Gamma(\gamma\alpha_{t-1}, \gamma\beta_{t-1})$. Therefore,

$$p(N_t|D^{t-1}) = \binom{\gamma\alpha_{t-1} + N_t - 1}{N_t}\Big(1 - \frac{1}{\gamma\beta_{t-1}+1}\Big)^{\gamma\alpha_{t-1}}\Big(\frac{1}{\gamma\beta_{t-1}+1}\Big)^{N_t}. \tag{1.2.10}$$

which is a negative binomial model denoted as

$$(N_t|D^{t-1}) \sim Negbin(r_t, p_t), \tag{1.2.11}$$

where $r_t = \gamma\alpha_{t-1}$ and $p_t = \frac{\gamma\beta_{t-1}}{\gamma\beta_{t-1}+1}$. Availability of one-step ahead predictive density in closed form is another advantage of he proposed model. Given (1.2.11), one can carry out one step ahead predictions and forecast interval calculations in a straightforward manner.

Although the k-step ahead predictive density is not analytically available, the k-step ahead predictive means can be easily obtained. Using a standard conditional expectation argument one can obtain $E(N_{t+k}|D^t)$ as follows

$$E(N_{t+k}|D^t) = E_{\theta_{t+k}}\{E(N_{t+k}|\theta_{t+k}, D^t)\} = E(\theta_{t+k}|D^t). \tag{1.2.12}$$

Furthermore, using the state equation we have

$$E(\theta_{t+k}|D^t) = E(\theta_t|D^t)\prod_{n=t+1}^{t+k}\frac{E(\epsilon_n|D^t)}{\gamma} = E(\theta_t|D^t) = \frac{\alpha_t}{\beta_t}, \tag{1.2.13}$$

where $E(\epsilon_n|D^t) = \gamma$ for any n. Therefore, combining (1.2.12) and (1.2.13), we can write

$$E(N_{t+k}|D^t) = E(\theta_{t+k}|D^t) = \frac{\alpha_t}{\beta_t}. \tag{1.2.14}$$

Due to the random walk type of structure introduced in (1.2.3), the above result simply indicates that k-step ahead forecasts given that we have observed counts up to time t are equal to α_t/β_t.

1.2.1 Learning about the discount parameter, γ

We can treat the discount factor γ as an unknown quantity and describe our uncertainty about it via a prior distribution, say $p(\gamma)$. Given D^t, the likelihood function of γ is given by

$$L(\gamma; D^t) = \prod_{i=1}^{t} p(N_i|D^{i-1}, \gamma), \tag{1.2.15}$$

where $p(N_i|D^{t-1}, \gamma)$ is negative binomial as in (1.2.11). The posterior distribution of γ can then be obtained as

$$p(\gamma|D^t) \propto \prod_{i=1}^{t} p(N_i|D^{i-1}, \gamma)p(\gamma). \tag{1.2.16}$$

For any choice of prior $p(\gamma)$ in (1.2.16) the posterior distribution can not be obtained analytically. However, we can always sample from the posterior γ using an MCMC method such as the Metropolis-Hastings algorithm. Alternatively, a discrete prior can be used for γ over $(0, 1)$. For example, a discrete uniform prior between 0.01 and 0.99 can be a reasonable choice and this will be considered in our examples.

1.2.2 Joint smoothing distribution of the Poisson rates

In addition to filtering and forecasting distributions that were obtained in the previous section, one can also obtain the smoothing distribution of the Poisson rate for retrospective type of analysis. In other words, given that we have observed the count data, D^t at time t, we will be interested in the distribution of $(\theta_{t-k}|D^t)$ for all $k \geq 1$.

We can write

$$p(\theta_{t-k}|D^t) = \int p(\theta_{t-k}|\theta_{t-k+1}, D^t)p(\theta_{t-k+1}|D^t)d\theta_{t-k+1}, \tag{1.2.17}$$

where $p(\theta_{t-k}|\theta_{t-k+1}, D^t)$ is obtained via the Bayes' rule as

$$\begin{aligned} p(\theta_{t-k}|\theta_{t-k+1}, D^t) &= \frac{p(\theta_{t-k}|\theta_{t-k+1}, D^{t-k})p(N^*|\theta_{t-k}, \theta_{t-k+1}, D^{t-k})}{p(N^*|\theta_{t-k+1}, D^{t-k})} \\ &= p(\theta_{t-k}|\theta_{t-k+1}, D^{t-k}), \end{aligned}$$

where $N^* = \{N_{t-k+1}, \cdots, N_t\}$. Here, given θ_{t-k+1}, N^* is independent of θ_{t-k}. In other words, $p(N^*|\theta_{t-k}, \theta_{t-k+1}, D^{t-k}) = p(N^*|\theta_{t-k+1}, D^{t-k})$. Thus, (1.2.17) reduces to

$$p(\theta_{t-k}|D^t) = \int p(\theta_{t-k}|\theta_{t-k+1}, D^{t-k})p(\theta_{t-k+1}|D^t)d\theta_{t-k+1}. \tag{1.2.18}$$

We can not obtain (1.2.18) analytically, however we can use Monte Carlo methods to draw samples from $p(\theta_{t-k}|D^t)$. This requires us to develop an efficient algorithm which would lead us to sample from the joint density, i.e $p(\theta_1, \cdots, \theta_t|\gamma, D^t)$, and then collect the samples corresponding to $p(\theta_{t-k}|\gamma, D^t)$ for all $k \geq 1$. Due to the Markovian nature of the state parameters, we can rewrite $p(\theta_1, \cdots, \theta_t|\gamma, D^t)$ as

$$p(\theta_t|\gamma, D^t)p(\theta_{t-1}|\theta_t, \gamma, D^{t-1}) \cdots p(\theta_1|\theta_2, \gamma, D^1). \tag{1.2.19}$$

We note that $p(\theta_t|\gamma, D^t)$ is available from (1.2.8) and $p(\theta_{t-1}|\theta_t, \gamma, D^{t-1})$ for any t can be obtained as follows

$$p(\theta_{t-1}|\theta_t, \gamma, D^{t-1}) \propto p(\theta_t|\theta_{t-1}, \gamma, D^{t-1})p(\theta_{t-1}|\gamma, D^{t-1}), \tag{1.2.20}$$

where the first term is available from (1.2.3) and the second term from (1.2.5). It would be straightforward to show that $(\theta_{t-1}|\theta_t, \gamma, D^{t-1}) \sim ShGamma[(1-\gamma)\alpha_{t-1}, \beta_{t-1}; (\gamma\theta_t, \infty)]$, that is a shifted gamma density defined over $\gamma\theta_t < \theta_{t-1} < \infty$. Therefore, given (1.2.19) and the posterior distribution of γ from (1.2.16), we can sample from $p(\theta_1, \cdots, \theta_t|\gamma, D^t)$ by sequentially simulating the individual Poisson rates as follows

1. Generate $\gamma^{(i)}$ from $p(\gamma|D^t)$.
2. Using the generated $\gamma^{(i)}$, sample $\theta_t^{(i)}$ from $(\theta_t|\gamma^{(i)}, D^t)$.
3. Using the generated $\gamma^{(i)}$, for each $n = t-1, \cdots, 1$ generate $\theta_n^{(i)}$ from $(\theta_n|\theta_{n+1}^{(i)}, \gamma, D^n)$.

If we repeat the above a large number of times, then we obtain samples from $p(\theta_1, \cdots, \theta_t|\gamma, D^t)$ which allows us to obtain a density estimate for $p(\theta_{t-k}|\gamma, N^{(t)})$ for all $k \geq 1$. The above simulation scheme is referred to as the forward filtering backward sampling (FFBS) algorithm as described in Fruhwirth-Schnatter (1994).

1.3 A Discrete Time Poisson Model with Covariates

It is possible to extend the basis model by considering the effects of covariates on the stochastic Poisson rate. As before let N_t be the number of occurrences of an event during time t and ν_t be its rate defined via

$$\nu_t = \theta_t e^{\boldsymbol{\psi}' \boldsymbol{z_t}}, \tag{1.3.1}$$

where $\boldsymbol{z_t}$ is the vector of the covariates and $\boldsymbol{\psi}$ is its parameter vector. Here, θ_t acts like a latent baseline rate which evolves over time but free of covariate effects. Given ν_t, we assume that number of occurrences of an event during time t is a modulated non-homogeneous Poisson process,

$$(N_t|\theta_t, \boldsymbol{\beta}, \boldsymbol{z_t}) \sim Pois(\theta_t e^{\boldsymbol{\psi}' \boldsymbol{z_t}}). \tag{1.3.2}$$

For the state evolution of the latent rate, θ_t we assume the same structure as before given by (1.2.2). In addition we assume that the conditional distribution of $(\theta_{t-1}|\boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1})$ follows a gamma density as

$$(\theta_{t-1}|\boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1}) \sim Gamma(\alpha_{t-1}, \beta_{t-1}). \tag{1.3.3}$$

Therefore, the conditional posterior density of θ_t given $\boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1}$ can be obtained via

$$p(\theta_t|\boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1}) = \int_{\gamma\theta_t}^{\infty} p(\theta_t|\theta_{t-1}, D^{t-1}) p(\theta_{t-1}|\boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1}) d\theta_{t-1}, \tag{1.3.4}$$

which reduces to a gamma density as

$$(\theta_t|\boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1}) \sim Gamma(\gamma\alpha_{t-1}, \gamma\beta_{t-1}). \tag{1.3.5}$$

Furthermore, the conditional filtering density of θ_t given $\boldsymbol{\psi}, \boldsymbol{z_t}, D^t$ can be obtained using (1.3.2) and (1.3.5) via the Bayes Rule which can be shown to be Gamma distributed as

$$(\theta_t|\boldsymbol{\psi}, \boldsymbol{z_t}, D^t) \sim Gamma(\alpha_t, \beta_t), \tag{1.3.6}$$

where $\alpha_t = \gamma\alpha_{t-1} + N_t$ and $\beta_t = \gamma\beta_{t-1} + (e^{\boldsymbol{\psi}'\boldsymbol{z_t}})$. Such an update scheme implies that as we learn more about the count process over time, we update our uncertainty about the Poisson rate as a function of both counts over time and of covariate effects via β_t as implied by (1.3.6).

The one-step ahead conditional predictive distribution at time t given $\boldsymbol{\psi}, \boldsymbol{z_t}$ and D^{t-1} can be shown to be negative binomial as

$$(N_t|D^{t-1}, \boldsymbol{\psi}, \boldsymbol{z_t}) \sim Negbin(r_t, p_t), \tag{1.3.7}$$

where $r_t = \gamma\alpha_{t-1}$ and $p_t = \frac{\gamma\beta_{t-1}}{\gamma\beta_{t-1}+e^{\boldsymbol{\psi}'\boldsymbol{z_t}}}$. (1.3.7) implies that given the covariates and the counts up to time $t-1$, forecast for time t is a function of the observed default counts up to time $t-1$ adjusted by the corresponding covariates. The conditional mean of $(N_t|D^{t-1}, \boldsymbol{\psi}, \boldsymbol{z_t})$, can be computed via

$$E(N_t|D^{t-1}, \boldsymbol{\psi}, \boldsymbol{z_t}) = \frac{\alpha_{t-1}}{\beta_{t-1}} e^{\boldsymbol{\psi}'\boldsymbol{z_t}}. \tag{1.3.8}$$

Since all conditional distributions previously introduced for the model with covariates are all dependent on the parameter vector $\boldsymbol{\psi}$, we need to discuss how to obtain the posterior density of $\boldsymbol{\psi}$ which can not be obtained in closed form, therefore we can use MCMC methods to generate samples of $\boldsymbol{\psi}$.

1.3.1 Markov chain Monte Carlo (MCMC) estimation

Our objective in this section is to obtain the posterior joint distribution of the model parameters given that we have observed all counts up to time t, that is $p(\theta_1, \cdots, \theta_t, \boldsymbol{\psi}|D^t)$. Since this joint distribution is not available in closed form we can use an MCMC method such as a Gibbs sampler to generate samples from it. In order to do so, we need to be able to generate samples from the full conditionals of $p(\theta_1, \cdots, \theta_t | \boldsymbol{\psi}, D^t)$ and $p(\boldsymbol{\psi} | \theta_1, \cdots, \theta_t, D^t)$, none of which are available as well known densities. Next we discuss how to generate samples from these densities.

The first full conditional, the conditional posterior distribution of $\boldsymbol{\psi}$ given the Poisson rates, $(\theta_1, \ldots, \theta_t)$, can be obtained via

$$p(\boldsymbol{\psi} | \theta_1, \cdots, \theta_t, z_t, D^t) \propto \prod_{i=1}^{t} \frac{exp\{\theta_i e^{\boldsymbol{\psi}' \boldsymbol{z_i}}\}(\theta_i e^{\boldsymbol{\psi}' \boldsymbol{z_i}})^{N_i}}{N_i!} p(\boldsymbol{\psi}), \tag{1.3.9}$$

where $p(\boldsymbol{\psi})$ is the prior for $\boldsymbol{\psi}$. Regardless of the prior selection for $\boldsymbol{\psi}$, (1.3.9) will not be a well known density. Therefore, we can use an MCMC algorithm such as the Metropolis Hastings to be able to generate samples from $p(\boldsymbol{\psi} | \theta_1, \cdots, \theta_t, z_t, D^t)$. Following Chib and Greenberg (1995), the steps in the Metropolis-Hastings algorithm can be summarized as follows

1. Assume the starting points $\boldsymbol{\psi^{(0)}}$ at $j = 0$.
   Repeat for $j > 0$,
2. Generate $\boldsymbol{\psi^*}$ from $q(\boldsymbol{\psi}^* | \boldsymbol{\psi^{(j)}})$ and u from $U(0, 1)$.
3. If $u \leq f(\boldsymbol{\psi}^{(j)}, \boldsymbol{\psi}^*)$ then set $\boldsymbol{\psi^{(j)}} = \boldsymbol{\psi}^*$; else set $\boldsymbol{\psi^{(j)}} = \boldsymbol{\psi^{(j)}}$ and $j = j + 1$,

where

$$f(\boldsymbol{\psi^{(j)}}, \boldsymbol{\psi}^*) = min \left\{ 1, \frac{\pi(\boldsymbol{\psi^*}) q(\boldsymbol{\psi^{(j)}} | \boldsymbol{\psi}^*)}{\pi(\boldsymbol{\psi^{(j)}}) q(\boldsymbol{\psi}^* | \boldsymbol{\psi^{(j)}})} \right\}. \tag{1.3.10}$$

In (1.3.10), $q(.|.)$ is the multivariate normal proposal density and $\pi(.)$ is given by (1.3.9) which is the density we need to generate samples from. If we repeat the above a large number of times then we obtain samples from $p(\boldsymbol{\psi} | \theta_1, \cdots, \theta_t, z_t, D^t)$.

Next we discuss how one can generate samples from the other full conditional distribution, $p(\theta_1, \cdots, \theta_t | \boldsymbol{\psi}, \boldsymbol{z_t}, D^t)$ using the FFBS algorithm. Due to the Markovian nature of the latent rates, using the chain rule we can rewrite the full conditional density as

$$p(\theta_t | \boldsymbol{\psi}, \boldsymbol{z_t}, D^t) p(\theta_{t-1} | \theta_t, \boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1}) \cdots p(\theta_1 | \theta_2, \boldsymbol{\psi}, \boldsymbol{z_t}, D^1). \tag{1.3.11}$$

We note that $p(\theta_t|\boldsymbol{\psi}, \boldsymbol{z_t}, D^t)$ is available from (1.3.6) and $p(\theta_{t-1}|\theta_t, \boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1})$ for any t can be obtained as follows

$$p(\theta_{t-1}|\theta_t, \boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1}) \propto p(\theta_t|\theta_{t-1}, \boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1})p(\theta_{t-1}|\boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1}). \tag{1.3.12}$$

It can be shown that $(\theta_{t-1}|\theta_t, \boldsymbol{\psi}, \boldsymbol{z_t}, D^{t-1}) \sim ShGamma[(1-\gamma)\alpha_{t-1}, \beta_{t-1}]$ where $\gamma\theta_t < \theta_{t-1} < \infty$, that is a shifted gamma density.

Therefore, given (1.3.11) and the posterior samples generated from the full conditional of $\boldsymbol{\psi}$, we can sample from $p(\theta_1, \cdots, \theta_t|\boldsymbol{\psi}, \boldsymbol{z_t}, D^t)$ by sequentially simulating the individual latent rates as follows

1. Assume the starting points $\theta_1^{(0)}, \cdots, \theta_t^{(0)}$ at $j = 0$.
   Repeat for $j > 0$,
2. Using the generated $\boldsymbol{\psi}^{(j)}$, sample $\theta_t^{(j)}$ from $(\theta_t|\boldsymbol{\psi}^{(j)}, \boldsymbol{z_t}, D^t)$.
3. Using the generated $\boldsymbol{\psi}^{(j)}$, for each $n = t-1, \cdots, 1$ generate $\theta_n^{(j)}$ from $(\theta_n|\theta_{n+1}^{(j)}, \boldsymbol{\psi}, \boldsymbol{z_t}, D^n)$ where $\theta_{n+1}^{(j)}$ is the value generated in the previous step.

If we repeat the above large number of times, then we obtain samples from the joint full conditional of the latent rates. Consequently, we can obtain samples from the joint density of the model parameters by iteratively sampling from the full conditionals, $p(\boldsymbol{\psi}|\theta_1, \cdots, \theta_t, z_t, D^t)$ and $p(\theta_1, \cdots, \theta_t|\boldsymbol{\psi}, \boldsymbol{z_t}, D^t)$, namely a full Gibbs sampler algorithm. The above can be extended to include the discount factor γ by extending the joint of $\boldsymbol{\psi}$ to include γ as $p(\gamma, \boldsymbol{\psi}|D^t, \boldsymbol{z_t})$ which can be sampled using the MCMC algorithm described previously.

1.4 Multivariate Extensions

It is possible to consider several extensions of the basis model to account for multivariate count data. For instance, the observations of interest can be the number of occurrences of an event during the day t of year j. Another possibility is to consider the analysis of J different

Poisson time series. For instance, for a given year, the weekly spending habits of J different households which can exhibit dependence can be modeled using such a structure. Several extensions have been proposed by Aktekin and Soyer (2011), where multiplicative Poisson rates for (1.2.2) are considered.

In what follows, we present a similar idea for J Poisson time series that are assumed to be affected by the same environment. We assume that

$$N_{jt} \sim Pois(\lambda_j \theta_t), \text{ for } j = 1, \ldots, J, \tag{1.4.1}$$

where λ_j is the arrival rate specific to the jth series and θ_t is the common term modulating λ_j. For example, in the case where N_{jt} is the number of grocery store trips of household j at time t, λ_j is the household specific rate and we can think of θ_t as the effect of common economic environment that the households are exposed to at time t. The values of $\theta_t > 1$ represents a more favorable economic environment than usual implying higher shopping rates.

This is analogous to the concept of an accelerated environment for operating conditions of components used by Lindley and Singpurwalla (1986) in life testing. Our case can be considered as a dynamic version of their set up since we have the Markovian evolution of θ_ts as

$$\theta_t = \frac{\theta_{t-1}}{\gamma} \epsilon_t, \tag{1.4.2}$$

where, as before, $(\epsilon_t | D^{t-1}, \lambda_1, \ldots, \lambda_J) \sim Beta[\gamma \alpha_{t-1}, (1-\gamma)\alpha_{t-1}]$ with $\alpha_{t-1} > 0$, $0 < \gamma < 1$, and $D^{t-1} = \{D^{t-2}, N_{1(t-1)}, \ldots, N_{J(t-1)}\}$. Furthermore, we assume that

$$\lambda_j \sim Gamma(a_j, b_j), \text{ for } j = 1, \ldots, J, \tag{1.4.3}$$

and a priori, λ_js are independent of each other as well as θ_0. Given θ_ts and λ_js, N_{jt}s are conditionally independent. In other words, all J series are affected by the same common environment and given that we know the uncertainty about the environment they will be independent.

At time 0, we assume that $(\theta_0 | D^0) \sim Gamma(\alpha_0, \beta_0)$, then by induction we can show that

$$(\theta_{t-1}|D^{t-1}, \lambda_1, \ldots, \lambda_J) \sim Gamma(\alpha_{t-1}, \beta_{t-1}), \tag{1.4.4}$$

and

$$(\theta_t|D^{t-1}, \lambda_1, \ldots, \lambda_J) \sim Gamma(\gamma\alpha_{t-1}, \gamma\beta_{t-1}). \tag{1.4.5}$$

In addition, the common filtering density at time t can be obtained via

$$(\theta_t|D^t, \lambda_1, \ldots, \lambda_J) \sim Gamma(\alpha_t, \beta_t), \tag{1.4.6}$$

where $\alpha_t = \gamma\alpha_{t-1} + N_{1t} + \ldots + N_{Jt}$ and $\beta_t = \gamma\beta_{t-1} + \lambda_1 + \ldots + \lambda_J$. Consequently, the marginal distributions of N_{jt} for any j can be obtained to be

$$p(N_{jt}|\lambda_j, D^{t-1}) = \binom{\gamma\alpha_{t-1} + N_{jt} - 1}{N_{jt}}\Big(1 - \frac{\lambda_j}{\gamma\beta_{t-1} + \lambda_j}\Big)^{\gamma\alpha_{t-1}}\Big(\frac{\lambda_j}{\gamma\beta_{t-1} + \lambda_j}\Big)^{N_{jt}}, \tag{1.4.7}$$

which is a negative binomial model as before. The multivariate distribution of $N_{1t}, \ldots, N_{Jt}$s can be obtained as

$$p(N_{1t}, \ldots, N_{Jt}|\lambda_1, \ldots, \lambda_J, D^{t-1}) = \frac{\Gamma(\gamma\alpha_{t-1} + \sum_j N_{jt})}{\Gamma(\gamma\alpha_{t-1})\prod_j \Gamma(N_{jt}+1)} \prod_j \left(\frac{\lambda_j}{\gamma\beta_{t-1} + \sum_j \lambda_j}\right)^{N_{jt}} \left(\frac{\gamma\beta_{t-1}}{\gamma\beta_{t-1} + \sum_j \lambda_j}\right)^{\gamma\alpha_{t-1}}, \tag{1.4.8}$$

which is a dynamic multivariate distribution of negative binomial type. The bivariate distribution $p(N_{it}, N_{jt}|\lambda_i, \lambda_j, D^{t-1})$ can be obtained as

$$\frac{\Gamma(\gamma\alpha_{t-1} + N_{it} + N_{jt})}{\Gamma(\gamma\alpha_{t-1})\Gamma(N_{it}+1)\,\Gamma(N_{jt}+1)}\Big(\frac{\gamma\beta_{t-1}}{\lambda_i + \lambda_j + \gamma\beta_{t-1}}\Big)^{\gamma\alpha_{t-1}}\Big(\frac{\lambda_i}{\lambda_i + \lambda_j + \gamma\beta_{t-1}}\Big)^{N_{it}}\Big(\frac{\lambda_j}{\lambda_i + \lambda_j + \gamma\beta_{t-1}}\Big)^{N_{jt}} \tag{1.4.9}$$

which is a bivariate negative binomial distribution for integer values of $\gamma\alpha_{t-1}$. The above distribution is the dynamic version of the negative binomial distribution proposed by Arbous and Kerrich (1951) for modeling number of accidents.

In addition the conditionals of N_{jt}s will also be negative binomial type distributions. The dynamic conditional mean (or regression) of N_{jt} given N_{jt} can be obtained as

$$E[N_{jt}|N_{it}, \lambda_i, \lambda_j, D^{t-1}] = \frac{\lambda_j(\gamma\alpha_{t-1} + N_{it})}{(\lambda_i + \gamma\beta_{t-1})}, \tag{1.4.10}$$

which is linear in N_{it}. It can be easily seen that the bivariate counts are positively correlated and the correlation is given by

$$Cor(N_{it}, N_{jt}|\lambda_i, \lambda_j, D^{t-1}) = \sqrt{\frac{\lambda_i \lambda_j}{(\lambda_i + \gamma\beta_{t-1})(\lambda_j + \gamma\beta_{t-1})}}. \quad (1.4.11)$$

Other properties of the dynamic multivariate distribution are given in Aktekin et al. (2014).

The estimation for the model using MCMC would be straightforward using the FFBS algorithm for θ_ts in conjunction with a Gibbs sampler step for the λ_js whose full conditionals are given by

$$p(\lambda_j|\theta_1, \ldots, \theta_t, D^t) \sim Gamma(a_{jt}, b_{jt}), \quad (1.4.12)$$

where $a_{jt} = a_j + N_{j1} + \ldots + N_{jt}$ and $b_{jt} = b_j + \theta_1 + \ldots + \theta_t$. Therefore, by iteratively sampling from the conditional distributions of $(\theta_1, \ldots, \theta_t|\lambda_1, \ldots, \lambda_J, D^t)$ using the FFBS algorithm and $(\lambda_j|\theta_1, \ldots, \theta_t, D^t)$ for all j, one can obtain samples from the full joint distribution of all model parameters, $(\theta_1, \ldots, \theta_t, \lambda_1, \ldots, \lambda_J|D^t)$.

1.5 Numerical Examples

In order to show how the models are applied to real count data, we used three data sets, time series counts of the number of calls arriving to a call center in a given time interval, number of people who defaulted in a given mortgage pool and the number of weekly grocery store visits for households. In the sequel, we discuss the implementation and the estimation of the proposed Poisson time series models using these three examples.

1.5.1 Example I: Call center arrival count data

To show the use of the basis model, we considered the time series of counts of call center arrivals during different intervals of 164 days from an anonymous U.S. commercial bank as discussed in Aktekin and Soyer (2011). Each day consists of 169 time intervals each of which is 5 minutes of duration. In a given day, the call center is operational between 7:00 AM and

9:05 PM. We only used the first week of the data for illustration purposes.

We consider each day of the week (Monday-Friday) separately and assume that the behavior of a given day is the same from one week to another. In other words we assume that call arrival process for Monday on any week will exhibit similar behavior to any other Monday in another week. We consider such a behavior by assuming the weekly call arrivals for any day are conditionally independent from one week to another, that is given the model parameters the weekly call arrivals for a given day are assumed to be independent. Such an approach would be of interest to call center managers who would like to be able to determine staff schedules in advance for different time intervals in a given day. For example, given call arrivals count data for Monday from the previous weeks, we can carry out inference for all the time periods for next Monday. Similarly if we have the call arrivals data for all other days, we can easily provide inference and forecasts for the whole next week.

Given the filtering of the latent rates, the discrete priors of γ for each day over $(0, 1)$ and their posterior estimation, we obtained the posterior means of the latent arrival rates for a particular time interval in a given day in the light of the whole data (that is 164 days with each having 169 time intervals). These are shown in Figure 1.1 from which a certain type of ordering between the days of the week can be inferred. As such, we set the initial prior parameters for the arrival rate in as $\alpha_0^i = \alpha_0$ and $\beta_0^i = \beta_0$ for all i with i representing a specific day of the week.

Figure 1 near here.

Furthermore, summary statistics for the posterior discounting behavior, γ for each day of the week are shown in Table 1.1. Discounting occurs on the sum of the previously observed values of the call arrivals for a given period, therefore it is a function of the dimension of data used. Each day seems to exhibit a slightly different discount behavior. The fact that the posterior means of the discounting terms are getting smaller as we observe more data, indicates that the model emphasizes arrival counts observed during the within-day interval of interest (say t) more than the previously observed arrival counts (say $t-1, \cdots, 1$).

Table 1 near here.

1.5.2 Example II: Mortgage default count data

In illustrating the use of the basis model and the model with covariates, we used data provided by Federal Housing Administration of the U.S. Department of Housing and Urban Development as analyzed in detail by Aktekin et al. (2013). In our analysis, we use 144 monthly defaulted FHA insured single-family 30-year fixed rate mortgage loans from 1994 in the Atlanta region. In addition, we make use of covariates such as the regional conventional mortgage home price index (CMHPI), federal cost of funds index (COFI), the homeowner mortgage financial obligations ratio (FOR) and regional unemployment rate (Unemp). A time series plot of the monthly mortgage count data under study is shown in Figure 1.2 where a non-stationary behavior that can be captured by our Poisson state space models is observed.

Figure 2 near here.

In analyzing the default count data, the discounting factor γ introduced in (1.2.2) was assumed to follow a discrete uniform distribution defined over $(0, 1)$ in order to keep the updating/filtering tractable. The posterior distribution of γ was obtained via (1.2.16) and is shown in the left panel of Figure 1.3. Thus, given the posterior of γ and the FFBS algorithm, it is possible to obtain the retrospective fit of counts. An overlay plot of the mean default rates and the actual data is shown in the right panel of Figure 1.3. The availability of the joint distribution of the default rate over time, i.e. $p(\theta_1, \cdots, \theta_t|D^t)$ would be of interest to institutions that are managing the loans for the purposes of risk management. Furthermore, Bayesian approach allows direct comparison of the Poisson rates (in this case default rates) during different time periods probabilistically. For instance, it would be straight forward to compute the probability that default rate during the second month is greater than that of the first month for a given cohort, i.e. $p(\theta_2 \geq \theta_1|D^{144})$.

Figure 3 near here.

In order to take into account the effects of covariates (macroeconomic variables in this case) on the default rate, we used the model with covariates. In doing so, we assumed the prior of

γ to be continuous uniform over $(0, 1)$ and the covariate coefficients, $\boldsymbol{\psi}$, to be independent normal distributions. The MCMC algorithm was run for 10,000 iterations with a burn-in period of 2,000 iterations with no convergence issues. The posterior density plots of $\boldsymbol{\psi}$ are shown in Figure 1.4 and of γ in Figure 1.5 which exhibits similar behavior to the posterior discounting term obtained for the basis model as in the left panel of Figure 1.3.

Figures 4-5 near here.

Table 1.2 shows the posterior summary statistics for the covariates. All macroeconomic variables seem to have fairly significant effects on the default rate. In summary, the regional conventional mortgage home price index (CMHPI), federal cost of funds index (COFI) and the regional unemployment rate (Unemp) have positive effects on default counts. For instance, as unemployment tends to go up, the model suggests that the number of people defaulting tend to increase for the cohort under study. On the other hand, the homeowner financial obligations ratio (FOR) seem to decrease the expected number of defaults as it goes up, namely as the burden of repayment becomes relatively easier then home owners are less likely to default.

Table 2 near here.

Figure 1.6 shows the fit of model with covariates to data from which the existence of a reasonably good fit can be inferred. Furthermore, the behavior of the latent default rates, θ_ts can be described via their joint distribution, that is $p(\theta_1, \cdots, \theta_{144}|D^{144})$. A boxplot of θ_ts is shown in Figure 1.7 which provides insights on the stochastic and temporal behavior of the latent rate given the count data at hand and the relevant covariates.

Figures 6-7 near here.

1.5.3 Example III: Household Spending Example

Our final example considers utilizes the multivariate extension of the basis model in the context of household spending. In order to illustrate the workings of the multivariate model in a simple setup, we considered a bivariate count data. However, let us emphasize that the multivariate count model can be applied to higher orders relatively easily. The data is regarding the weekly grocery store visits of 540 Chicago based households accumulated over 104 weeks, from which we considered 2 households. In other words, we have 2 different Poisson time series for different households and assume that their visits to the grocery store can be modeled by 1.4.1. Our assumption is that each household's visit to the grocery store is affected by the same environment, i.e. the economic situation, weather and so on. Namely, we assume that the grocery store arrival process of a household in Chicago, will exhibit some common behavior to that of another household.

Figure 8 near here.

This approach would be of interest to grocery store managers who would like to be able to differentiate the common effect from individual effects for store promotion purposes. For example, analysis and inferences from θ_t may allow managers to carry out store-wide promotion activities, whereas analysis on individual λ_js will allow managers to target specific households to promote store visits. For illustrative purposes, we fixed the discount factor at $\gamma = 0.5$, and set the initial prior parameters as $\alpha_0^j = \alpha_0$ and $\beta_0^j = \beta_0$ for both js with j representing each individual household. As before, the behavior of the common arrival rates, θ_ts can be described via their joint distribution, $p(\theta_1, \cdots, \theta_{104}|D^{104})$. We present a boxplot of θ_ts in Figure 1.9 which provides insights on the temporal behavior of the common rate given the count data at hand and the individual rate λ_js. Specifically, we find a drop in the common rates in weeks 29-35 and in weeks 79-85 which indicates a possible seasonal effect occurring over the calender year. Note that, as previously discussed, such seasonal effects can be easily incorporated into the model as covariates.

Figure 9 near here.

Furthermore, the posterior density plots of $\boldsymbol{\lambda_j}$s are shown in Figure 1.10. Clearly household 1 can be characterized by a higher rate compared to household 2. Store manager may utilize our findings here in multiple ways. First, the manager may identify the season associated with low common arrival rates and run a store-wide promotion strategy to lure customers in. In addition, store managers faced with limited budget may also wish to focus their promotion efforts on household 2 with a lower arrival rate rather than household 1. Finally, one may also extend the model to include covariates to identify the reasons behind differences in individual household λ_js.

Figure 10 near here.

1.6 Conclusion

In this study, we introduced a general class of Poisson time series models. In doing so, we first discussed univariate discrete time state space models with Poisson measurements and their Bayesian inference via MCMC methods. Furthermore, we discussed issues of sequential updating, filtering, smoothing and forecasting. We also introduced modeling strategies for multivariate extensions. In order to show the implementation of the proposed models, we used real count data from different disciples such as finance, operations and marketing.

We believe that several future directions can be pursued as a consequence of this study. One such area is to treat γ as a time varying or a series specific discount factor which could potentially create challenges in parameter estimation. Another possibility for estimation purposes is to investigate the implementation of sequential particle filtering methods instead of MCMC that are widely used for state space applications.

Bibliography

Aktekin, T., Kim, B., and Soyer, R. (2014). Dynamic multivariate distributions for count data: A bayesian approach. Technical report, Institute for Integrating Statistics in Decision Sciences, The George Washington University.

Aktekin, T. and Soyer, R. (2011). Call center arrival modeling: A Bayesian state-space approach. *Naval Research Logistics*, 58(1):28–42.

Aktekin, T., Soyer, R., and Xu, F. (2013). Assessment of mortgage default risk via Bayesian state space models. *Annals of Applied Statistics*, 7(3):1450–1473.

Arbous, A. G. and Kerrich, J. (1951). Accident statistics and the concept of accident proneness. *Biometrics*, 7:340–432.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hasting algorithm. *The American Statistician*, 49(4):327–335.

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. and Winkelmann, R. (2001). Markov chain monte carlo analysis of correlated count data. *Journal of Business and Economic Statistics*, 19(4):428–435.

Davis, R. A., Dunsmuir, W. T. M., and Streett, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika*, 90(4):777–790.

Davis, R. A., Dunsmuir, W. T. M., and Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*, 87(3):491–505.

Durbin, J. and Koopman, S. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society. Series B*, 62(1):3–56.

Freeland, R. K. and McCabe, B. P. M. (2004). Analysis of low count time series data by Poisson autocorrelation. *Journal of Time Series Analysis*, 25(5):701–722.

Fruehwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika*, 93(4):827–841.

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

Harvey, A. C. and Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business and Economic Statistics*, 7(4):407–417.

Lindley, D. V. and Singpurwalla, N. D. (1986). Multivariate distributions for the life lengths of components of a system sharing a common environment. *Journal of Applied Probability*, 23:418–431.

Santos, T. R. D., Gamerman, D., and Franco, G. C. (2013). A non-Gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis*, 34(6):625–645.

Smith, A. F. M. and Gelman, A. E. (1992). Bayesian statistics without tears: A Sampling perspective. *The American Statistician*, 46(2):84–88.

Smith, R. and Miller, J. (1986). A non-Gaussian state space model and application to prediction of records. *Journal of the Royal Statistical Society. Series B*, 48(1):79–88.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, 75(4):621–629.

Table 1.1: Posterior means and standard deviations of γ for different days

Day	Mean	St.Dev
Mondays	.066	.0025
Tuesdays	.046	.0016
Wednesdays	.092	.0042
Thursdays	.084	.0039
Fridays	.075	.0032

Table 1.2: Posterior statistics for $\boldsymbol{\psi}$ and γ for the model with covariates

Statistics	ψ_{CMHPI}	ψ_{COFI}	ψ_{FOR}	ψ_{Unemp}	γ
25th	0.0063	0.7003	-1.5430	0.6252	0.2281
Mean	0.0160	0.8717	-1.3002	0.8191	0.2466
75th	0.0256	1.0510	-1.0550	1.0117	0.2643
St.Dev	0.0141	0.2663	0.3606	0.2826	0.0270

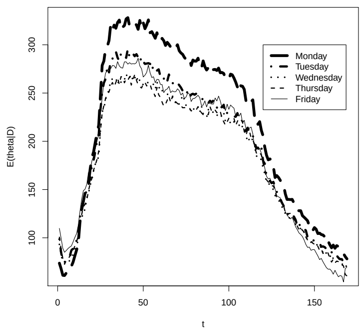

Figure 1.1: Posterior arrival rates for different days of the week

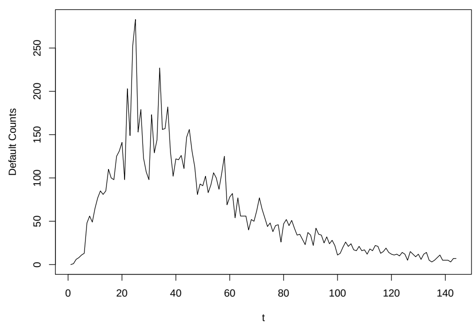

Figure 1.2: Time series plot of monthly default counts

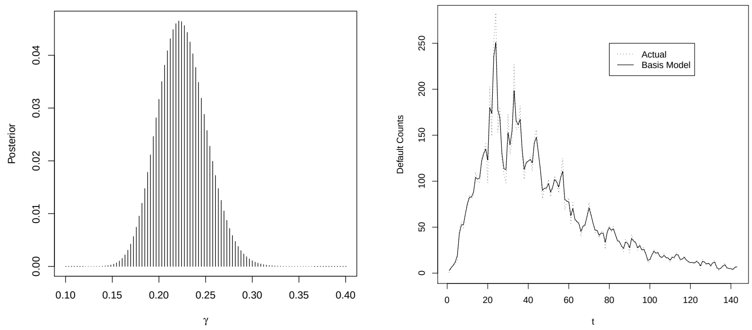

Figure 1.3: Posterior γ (left) and the retrospective fit to count data (right)

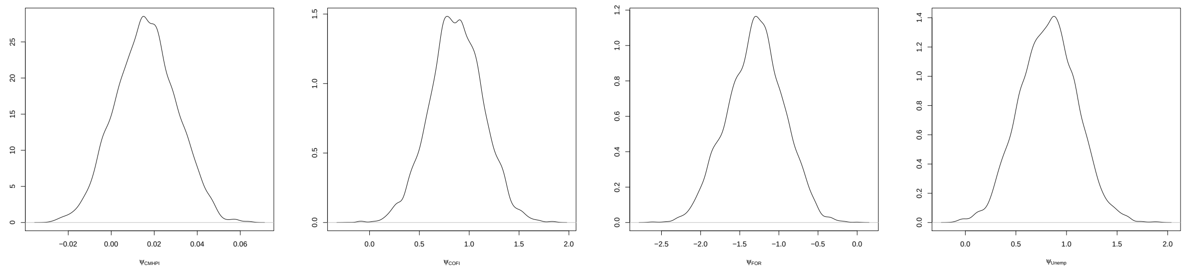

Figure 1.4: Posterior density plots of $\boldsymbol{\psi}$

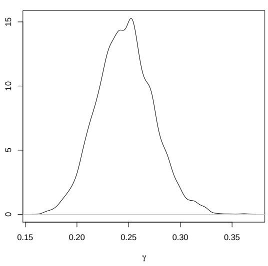

Figure 1.5: Posterior density plot of γ for the model with covariates

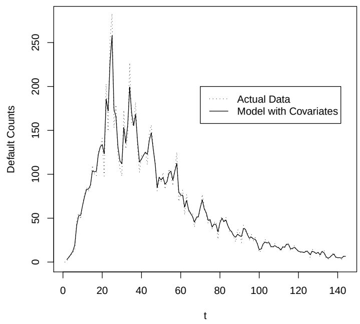

Figure 1.6: Retrospective fit of the model with covariates to data

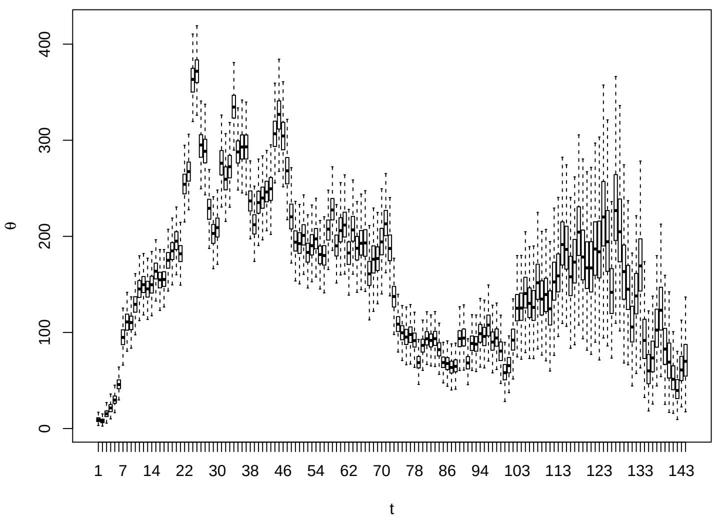

Figure 1.7: Boxplots for the latent rates, θ_ts from $p(\theta_1, \cdots, \theta_{144}|D^{144})$

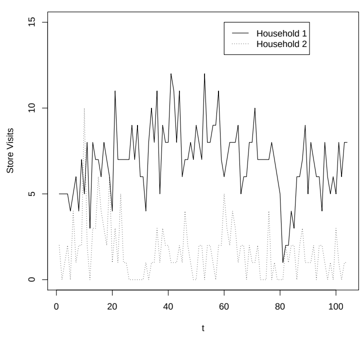

Figure 1.8: Time series plot of weekly grocery store visits

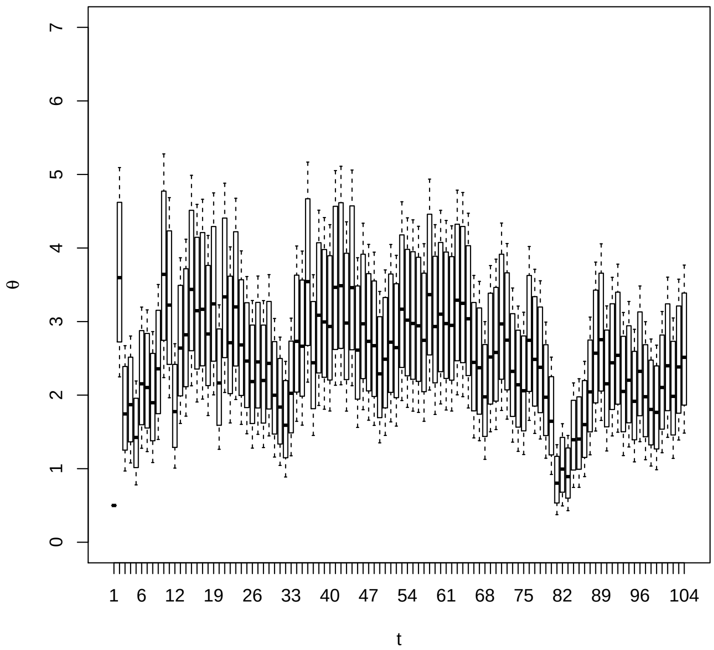

Figure 1.9: Boxplots for the common rates, θ_ts from $p(\theta_1, \cdots, \theta_{104}|D^{104})$

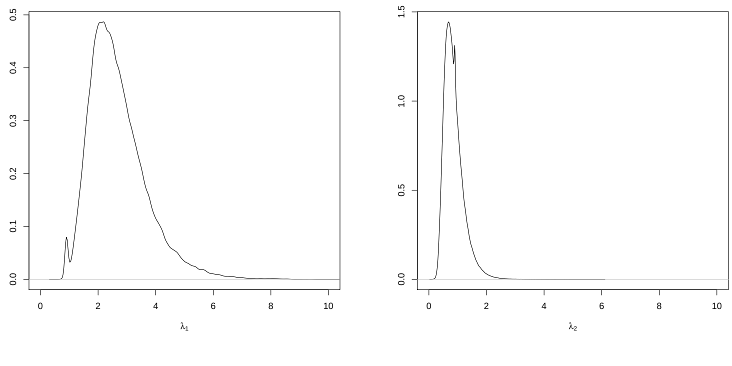

Figure 1.10: Posterior density plots of λ_1 and λ_2