2-7-2019

Zero-inflated, hurdle and bivariate parameter-driven count models

Huda Al-Wahsh
University of Windsor

ZERO-INFLATED, HURDLE AND BIVARIATE PARAMETER-DRIVEN COUNT MODELS

by

Huda Al-Wahsh

A Dissertation

Submitted to the Faculty of Graduate Studies

through the Department of Mathematics and Statistics

in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy at the

University of Windsor

Windsor, Ontario, Canada

2019

ZERO-INFLATED, HURDLE AND BIVARIATE PARAMETER-DRIVEN COUNT MODELS

by

Huda Al-Wahsh

APPROVED BY:

M. Ghahramani, External Examiner
University of Winnipeg

R. Kent
School of Computer Science

M. Hlynka
Department of Mathematics and Statistics

S. Paul
Department of Mathematics and Statistics

A. Hussein, Advisor
Department of Mathematics and Statistics

February 5, 2019

Author's Declaration of Originality

I hereby declare that this dissertation incorporates the outcome of joint research undertaken in collaboration with my supervisor Professor A. Hussein. I also declare that no part of this dissertation has been published.

I certify that, to the best of my knowledge, my dissertation does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my dissertation, published or otherwise, are fully acknowledged in accordance with standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained written permission from the copyright owner(s) to include such material(s) in my dissertation and have included copies of such copyright clearances to the appendix.

I declare that this is a true copy of my dissertation, including any final revisions, as approved by my dissertation committee and the Graduate Studies office, and that this dissertation has not been submitted for a higher degree to any other University or Institution.

Abstract

A time series is a collection of observations made sequentially through time. Examples occur in a variety of fields, ranging from medicine to engineering. The analysis of time series of counts is one of the rapidly developing areas in time series modeling. In time series, it is unlikely that neighbouring observations are independent. To accommodate potential correlation for count data, two main classes of models are frequent in the literature: parameter-driven and observation-driven models. Central to both classes are the generalized linear models (GLMs). Parameter-driven models result when temporal random effects are used in the GLM to accommodate the autocorrelations.

In this dissertation we propose zero-inflated and hurdle specifications for both Poisson and negative binomial parameter-driven models. We employ the data cloning approach as the numerical tool for performing inferences about the models. We carry out intensive simulations to examine the performance of the proposed methodologies. An application of the methods to a data set on the daily counts of emergency department visits for asthma cases in Ontario, Canada, is also provided.

The second focus of this dissertation is to model dependence in bivariate time series of counts. In this direction, we propose two parameter-driven models based on a commonly used bivariate Poisson specification. The first model employs one latent process through the cross-correlation parameter of the bivariate Poisson distribution, thus leading to common temporal autocorrelations between the components of the bivariate Poisson, while the second model uses two latent processes to introduce separate autocorrelations in the two marginal processes. An intensive simulation study and real data applications are also provided in these scenarios.

Dedication

Dedicated to the Memory of My Mother.

Acknowledgements

All glorification and gratitude to Almighty who has given me the ability to accomplish this dissertation.

I would like to express my deepest gratitude to my supervisor, Dr. A. Hussein for his guidance, help, continuous encouragement and financial support. This dissertation would not have been completed without his continuous academic advice.

I would like to thank the advisory committee members, Dr. M. Hlynka, Dr. S. R. Paul and Dr. R. Kent for reading my dissertation, and offering constructive suggestions. My great gratitude also goes to the external examiner Dr. M. Ghahramani from University of Winnipeg for her valuable comments and advice.

My sincere thanks also goes to my friend Khiria. Your help and kindness made my life in Windsor wonderful.

Last but not least, I would like to thank my father and my brother Ismail. Your love and expectation motivate me to be brave and step forward. I love you, I am extremely grateful for what you have done for me and I am here because of you!

Table of Contents

List of Tables

Page

List of Figures

List of Abbreviations

AIC Akaike information criteria

AR(p) Autoregressive process of order p

BP1 Bivariate Poisson with one latent process

BP2 Bivariate Poisson with two latent processes

DC Data cloning

MCMC Markov chain Monte-Carlo

MLE Maximum likelihood estimator

NPARH Negative binomial autoregressive hurdle

NPH Negative binomial hurdle

PARH Poisson autoregressive hurdle

PH Poisson hurdle

ZINB Zero-inflated negative binomial

ZIP Zero-inflated Poisson

Chapter 1

Introduction and preliminaries

1.1 Introduction

A time series is a collection of observations made sequentially through time. Examples occur in a variety of fields, ranging from medicine to engineering. With time series, it is unlikely that neighbouring observations are independent.

To accommodate potential correlation for count data, Brockwell and Davis [7] and Davis et al. [18] described the so-called generalized state space models for non-Gaussian time series. A generalized state space model for a time series, consists of an observation variable and a state variable. The model is expressed in terms of conditional probability distributions for the observation given the state variables. Cox [14] characterized such models as observation-driven, and parameter-driven models. The observation specification is the same for both models. For observation-driven models the state equation depends on past observations, as well as other covariates. Generalized linear autoregressive moving average (GLARMA) and integer-valued GARCH models are examples, see Davis et al. [17] and Ferland et al. [23]. On the other hand, in parameter-driven models, the state equation commonly consists of a regression component and a latent process that cannot be

observed directly and which evolves independently of past and present values of the observed responses, see Zeger [67] and Jørgensen et al. [34] for examples.

The analysis of time series of counts, motivated by applications in various fields, is one of the rapidly developing areas in time series modeling. In recent years, there has been a considerable development of models for non-Gaussian time series. Generalized linear models (McCullagh and Nedler, [45]) are widely used for analyzing counts and other types of discrete data. The Poisson model provides the main instrument for modeling count data. To accommodate overdispersion, many researchers have turned to overdispersed Poisson and negative binomial regression models. The negative binomial distribution has flexibility in its parametrization and has been used differently by different authors. Hilbe [32] presented a detailed analysis of the nature and scope of the varieties of negative binomial model.

Count data with numerous zeros are common in a number of industrial applications. Lambert [41] proposed the zero-inflated Poisson (ZIP) regression model with an application to defects in manufacturing. An adaptation of Lambert's ZIP regression to the situation in which the response is an upper-bounded count was done by Hall [30] by proposing a zero-inflated binomial model. Furthermore, Ridout et al. [57] provided a score test for testing ZIP regression models against zero-inflated negative binomial alternatives. Ghosh et al. [27] presented a Bayesian analysis for a class of zero-inflated models which includes the ZIP models, as special cases. Also, Famoye and Singh [22] described a zero-inflated generalized Poisson regression to model domestic violence data.

Hurdle models provide an alternative way to model count data with excess zeroes. The hurdle model, proposed by Mullahy [49], is a two part model in which the two separate processes generating zeroes and positive counts are not constrained to be the same. The first part is a binary response that estimates

the probability that the zero hurdle (threshold) is crossed. The second part uses a truncated model to explain the observations above the hurdle. (Cameron and Trivedi, [10]). Theoretically the threshold could be any value, but it's usually taken at zero because this is most often meaningful in the context of the study objectives. Dobbie and Welsh [20] extended the hurdle model to take account of possible serial dependence between repeated observations and used generalized estimating equations to estimate the model parameters, while Min and Agresti [48] introduced a hurdle model with random effects for clustered and correlated counts.

The hurdle model is used in a variety of applications, for example in public health Dalrymple et al. [15] used hurdle models to accommodate extra zeros and heterogeneity found in the sudden infant death syndrome series. Rose et al. [59] used zero-inflated and hurdle models for modelling vaccine adverse event count data. In finance, Boucher et al. [5] used correlated random effects for hurdle models to model a panel of claim count data. For terrorism modelling, Porter [54] preferred the hurdle model to the zero-inflated model which assumes that the extra zeros are due to censoring while the hurdle model assumes that the extra zeros are due to a separate process, which must be overcome before the number of corresponding incidents are determined.

A major part of this dissertation is dedicated to the development of zero-inflated and hurdle parameter-driven models to accommodate correlations in time series of counts with excess zeros. We propose the use of the data cloning method to estimate the parameters of these models. Real data examples and numerical studies are presented.

Bivariate time series of counts arise in many applications where two counts are correlated and joint estimation is required. For example, such data occur in

epidemiology when the number of hospital admissions and the number of hospital discharges are examined, also when two diseases are observed and likely to be inter-dependent, in marketing when the number of firms which enter and exit an industry are studied across time.

The bivariate Poisson distribution is probably the most well-known bivariate discrete distribution, though its definition is not unique. For an excellent discussion of the various situations in which the distribution arises, we refer to Kocherlakota and Kocherlakota [38] and Johnson et al. [33].

The first-order integer-valued autoregressive (INAR(1)) model for pure time series was introduced by McKenzie [47] and later discussed by Al-Osh and Alzaid [1]. Brännäs [6] extended the model to account for explanatory variables. A large number of studies have considered the modeling of bivariate or multivariate count data assuming underlying Poisson distributions. For example, Quoreshi [55] proposed a bivariate integer-valued moving average (BINMA) model to fit bivariate time series of count data that are generated from stock transactions. Pedeli and Karlis [52] defined a bivariate INAR process and discussed alternative methods for the estimation of its unknown parameters. Liu [44] formulated a bivariate Poisson integer-valued GARCH model and demonstrated its stability properties.

All the previous models are classified as observation-driven models. Our second aim in this dissertation is to propose a parameter-driven model that is capable of modeling overdispersion and serial dependence between two time series of counts. In Chapter 4 we propose two parameter-driven models of bivariate Poisson, the first model with one latent process added to the cross-correlation parameter, and the second model with two latent processes to propose different correlation in the two time series.

Here is a list of the novel contributions in this thesis:

1. In Chapters 2 and 3:

   (a) We computed the moments of the Zero-inflated Poisson (ZIP) and the Zero-inflated Negative binomial (ZINB) parameter-driven models under the assumption of autoregressive latent process of order p.

   (b) We formulated the two models in a hierarchical format, appropriate for Data Cloning (DC) algorithm to be applied, then we carried out inferences based on MLEs obtained through the DC method.

2. In Chapter 4:

   (a) We proposed new hurdle parameter-driven models for both Poisson and negative binomial distributions with the help of an autoregressive Gaussian latent process of order p.

   (b) We formulated the two models in a hierarchical format, appropriate for the DC algorithm and then carried out inferences based on the MLEs obtained via the DC.

3. In Chapter 5:

   (a) The following two new parameter-driven bivariate Poisson models were introduced:

   1- BP1 model by including an AR(p) process to the cross correlation parameter of a bivariate Poisson distribution.

   2- BP2 model by including two latent processes, AR(p) and AR(q), in the marginal distributions of a bivariate Poisson model.

   These two models are useful, in situations where the components of a bivariate count time series have same temporal autocorrelation behavior

or different temporal autocorrelation behavior.

(b) We derived the moments of these new models, formulated them in hierarchical specifications and used the DC method to obtain the MLEs of their parameters.

According to the above layout of contributions, the rest of this chapter is organized as follows: basic introduction to Bayesian inference, Markov chain Monte Carlo methods, data cloning procedure for computing maximum likelihood estimates for hierarchical models, and an illustration of the difference of Akaike information criteria for model selection in the framework of hierarchical models.

In Chapter 2, we present the zero-inflated Poisson parameter-driven model and its parameter estimation via the data cloning method, results of some numerical simulation studies are reported and an illustrative example using the asthma dataset is given.

Chapter 3 presents the zero-inflated negative binomial parameter-driven model and its parameter estimation illustrated by some numerical simulations and the asthma dataset application.

In Chapter 4, we propose hurdle parameter-driven Poisson and negative binomial models and we estimate their parameters via data the cloning approach. To motivate both theoretical and methodological developments given in the chapter, real data sets are used for illustration.

In Chapter 5, we propose two parameter-driven models by using the bivariate Poisson distribution. A simulation study is conducted and applications on real data are presented. A summary and some future research is given in Chapter 6.

1.2 Bayes statistics and MCMC methods

In Bayesian inference there is a fundamental distinction between observable quantities y, that is the data, and unknown quantities θ. The unknown quantities could be statistical parameters, missing data, latent process,.... In Bayesian framework θ are treated as random variables, so we need probability statements about θ. The prior distribution expresses our uncertainty about θ before seeing the data, while the posterior distribution expresses our uncertainty about θ after seeing the data.

Markov Chain Monte-Carlo (MCMC) is a popular method for obtaining information about distributions, especially for estimating posterior distributions in Bayesian inference. It allows one to characterize a distribution without knowing all of the distribution's mathematical properties by randomly sampling values out of the distribution. Bayesian data analysis and MCMC techniques tremendous increase in popularity over the last decade is due to an increase in computational power which has made it affordable to do such computations.

The name MCMC combines two properties: Monte-Carlo and Markov chain. Monte-Carlo is the practice of estimating the properties of a distribution by examining random samples from the distribution. For example, instead of finding the mean of a specific distribution by calculating it directly from the distribution, a Monte-Carlo approach would draw a large number of random samples from this distribution and calculate the sample mean of those. Of course, calculating the mean of a large sample of numbers can be much easier than calculating the mean directly from the distribution, especially when random samples are easy to draw, and when the distribution is hard to work with in other ways. The Markov chain property of MCMC is the idea that the random samples are generated by a spe-

cial sequential process. Each random sample is used to generate the next random sample, hence producing the chain. Each new sample in the chain depends on the one before it and does not depend on any samples before the previous one, this is the "Markov" property.

In a Bayesian approach, the chain of values produced converges to its equilibrium distribution which is the joint posterior distribution. The theory of how to construct this chain to achieve the proper distribution can be quite complicated, but suffice it to say that there are some general methods that can be used in most problems and that are implemented in available software(Roberts and Rosenthal [58]). In this dissertation we used one of the most prevalent kinds of software, JAGS.

JAGS is a program for analysis of Bayesian hierarchical models using MCMC computations; it designed to work closely with the R language. We used *coda* package to analyse the output and *rjags* package to work directly with JAGS from within R.

Once the MCMC chain has been run and simulated samples from the algorithm have been stored, we need to perform some diagnostics on the simulations to determine if they approximately represent the posterior distribution of interest. There are few relatively simple diagnostics of algorithm convergence:

1. Monitoring the trace plot of the parameter samples: Once convergence has been reached, samples should look like a random scatter about a stable mean value, there should be no obvious trend or change in spread.

2. The plot of autocorrelation function of parameter samples: we would expect the kth lag autocorrelation to be smaller as k increases. If autocorrelation is still relatively high for higher values of k, this indicates high degree of

correlation between draws and slow mixing.

3. Multivariate $\widehat{R}$ values for MCMC chain convergence: The $\widehat{R}$ statistic measures the ratio of the average variance of samples within each chain to the variance of the pooled samples across chains; if all chains are at equilibrium, these will be the same and $\widehat{R}$ will be close to one. If the chains have not converged to a common distribution, the $\widehat{R}$ statistic will be greater than one. (Brooks and Gelman [8])

1.3 Theory of data cloning

We use the data cloning method in order to obtain the maximum likelihood estimates of the parameters, $\widehat{\theta}$. Data cloning (DC) is a statistical computing method introduced by *Lele et al.* [42]. It exploits the computational simplicity of the Markov chain Monte Carlo (MCMC) algorithms used in the Bayesian statistical framework, to obtain the maximum likelihood point estimates and their standard errors for complex hierarchical models. The use of the data cloning algorithm is especially valuable for complex models, where the number of unknowns increases with sample size (*i.e.* with latent variables), because inference and prediction procedures are often hard to implement in such situations.

Consider the following form of a hierarchical model:

$$\begin{aligned} &\text{Hierarchy 1:} \quad Y = y|X = x \sim f(y|x, \theta_1), \\ &\text{Hierarchy 2:} \quad X \sim g(x|\theta_2), \end{aligned} \tag{1.3.1}$$

where y are observed and x are unobserved. The parameters of interest are

$\theta = (\theta_1, \theta_2)$ and the likelihood function for this hierarchical model is given by:

$$L(\theta, y) = \int f(y|x, \theta_1)\ g(x|\theta_2)\ dx.$$

We assume that the parameters are identifiable and that there is a unique mode (but possibly multiple smaller peaks) to the likelihood function.

To understand the idea of DC method, imagine a hypothetical situation where an experiment is repeated by K different observers, and all K experiments happen to result in exactly the same set of observations $y^{(K)} = (y, \cdots, y)$. The likelihood function based on the combination of the data from these K independent experiments is given by $L(\theta, y^{(K)}) = [L(\theta, y)]^K$. Notice two important features of this likelihood function:

(i) The location of the maximum of this function is exactly equal to the location of the maximum of $L(\theta, y)$.

(ii) The Fisher information matrix based on $L(\theta, y)$ times K equals the Fisher information matrix of this likelihood function.

It is easy to see that the posterior distribution of θ conditional on the data $y^{(K)} = (y, \cdots, y)$ is given by

$$\begin{aligned}\pi_K(\theta|y) &= \frac{\left[\int f(y|x, \theta_1) g(x|\theta_2) dx\right]^K \pi(\theta)}{C(K, y)} \\ &= \frac{[L(\theta, y)]^K \pi(\theta)}{C(K, y)},\end{aligned}$$

where

$$C(K, y) = \int \left[\int f(y|x, \theta_1) g(x|\theta_2) dx\right]^K \pi(\theta) d\theta$$

is the normalizing constant, and $\pi(\theta)$ is the prior distribution of the parameters.

Lele et al. [42] and [43] proved that as K becomes large, $\pi_K(\theta|y)$ converges to a multivariate normal distribution with mean equal to the MLE $\widehat{\theta}$ and variance-covariance matrix equal to $\frac{1}{K}I^{-1}(\widehat{\theta})$ where $I(\widehat{\theta})$ is the information matrix corressponding to the original likelihood function $L(\theta, y)$. This convergence is deterministic convergence of a sequence of functions and not the probabilistic convergence used in *Walker* [66]. Thus, the asymptotic variance of the ML estimate can be estimated by multiplying K times the variance of the kth cloned posterior distribution. One major advantage of the data cloning method is the invariance of the results to the choice of priors.

It follows then that if we can generate random variates $\theta_1, \theta_2, \cdots, \theta_B$ from $\pi_K(\theta|y)$ distribution, then we can use their mean and variance to obtain the MLE $\widehat{\theta}$ and its asymptotic varaince. Fortunately, such generation of random variates from $\pi_K(\theta|y)$ is quite easy using the MCMC algorithms. Determining the number of clones K is possible through disgnostics measures (*Lele et al.* [43]). These measures include:

1- Calculating the largest eigenvalue of the posterior variance covariance matrix. If the parameters are identifiable, then this measure should converge to 0 at a rate $\frac{1}{K}$.

2- Calculating mean square error, $\omega = \frac{1}{B}\sum_{i=1}^{B}(O_i - E_i)^2$, where E_i are the quantiles for χ^2_p random variable and $O_i = (\theta_i - \overline{\theta})^T V^{-1}(\theta_i - \overline{\theta})$.

3- Calculating correlation-like fit statistic, $r^2 = 1 - corr^2(O_i, E_i)$, where $corr^2(O_i, E_i)$ is Pearson's correlation. If this statistic and the one before are close to zero, it indicates that the $(\theta_i - \overline{\theta})^T V^{-1}(\theta_i - \overline{\theta}) \backsim \chi^2_p$ approximation is reasonable.

These measures and multivariate $\widehat{R}$ for MCMC chain convergence are available in *dclone* package in R software (*Sölymos*, [60]).

1.4 Model selection

Akaike Information Criteria (AIC) compares a set of statistical models to each other, a good model is the one that has minimum AIC among all the models. The formula for AIC is

$$AIC = 2d - 2ln(L(\widehat{\theta}, y)),$$

where d is the number of model parameters.

To compute AIC we need the maximized likelihood values which are not directly available for parameter-driven models. To overcome this limitation (*Ponciano et al.* [53]) used the complete likelihood function to compute AIC difference and used it to compare between two nested models. To illustrate, suppose that model 1 and model 2 are any two nested models, one can write

$$AIC_1 - AIC_2 = -2ln\left(\frac{L(\theta_1^{(1)}, \theta_2^{(1)}, y)}{L(\theta_1^{(2)}, \theta_2^{(2)}, y)}\right) + 2(d_1 - d_2),$$

where d_1 and d_2 are the number of estimated parameters under model 1 and 2, respectively.

Recall that the likelihood of a hierarchical model defined as in Equation [1.3.1] can be written as

$$L(\theta_1, \theta_2, y) = \int f(y|x, \theta_1)g(x|\theta_2)dx,$$

where y is a vector of observations and x is the vector of latent variables. The desired likelihood ratio evaluated at two different sets of parameter values can be estimated as:

$$\frac{L(\theta_1^{(1)}, \theta_2^{(1)}, y)}{L(\theta_1^{(2)}, \theta_2^{(2)}, y)} = \frac{1}{m} \sum_{i=1}^{m} \frac{f(y|x^{(i)}, \theta_1^{(1)}) g(x^{(i)}|\theta_2^{(1)})}{f(y|x^{(i)}, \theta_1^{(2)}) g(x^{(i)}|\theta_2^{(2)})},$$

where $x^{(1)}, x^{(2)}, \cdots, x^{(m)}$ are generated samples from the conditional distribution

$$h(x|y, \theta_1^{(2)}, \theta_2^{(2)}) \propto f(y|x, \theta_1^{(2)}) g(x|\theta_2^{(2)}).$$

For a collection of more than two models, the likelihood ratios need to be calculated for all pairs of models.

AIC differences greater than 2 are generally thought to be significant, and differences greater than 3 very significant. (*Burnham and Anderson*, [9]).

Chapter 2

Zero-inflated Poisson parameter-driven model

This chapter proceeds as follows: in Section 1 we present the Poisson parameter-driven model and some of its basic properties. Also, to accommodate potential correlation for count data with excess zeros, we propose a nonstationary zero-inflated Poisson parameter-driven model. In Section 2, the maximum likelihood estimators of the model parameters are obtained via the data cloning method. A simulation study is conducted in Section 3 and in Section 4, the techniques developed in earlier sections are applied to real data sets.

2.1 Poisson regression models

Poisson regression model is a form of a generalized linear model where the response variable is modelled as having a Poisson distribution and it is a natural choice when the response variable is an integer.

Let $\{Y_t : t = 1, 2, \cdots, n\}$ be a time series of observed counts, $x_t^T = (x_{t1}, \cdots, x_{tk})$ is the tth row of covariate matrix X and $\beta = (\beta_1, \cdots, \beta_k)^T$ are unknown k-

dimensional column vector of parameters. Then a Poisson regression model is given as follows:

$$f(y_t|x_t) = \frac{e^{-\lambda_t}\lambda_t^{y_t}}{y_t!}, \quad y_t = 0, 1, 2, \cdots$$

where

$$log(\lambda_t) = x_t^T\beta = \sum_{j=1}^{k} \beta_j x_{tj}.$$

For count data of independent observations with excess zeros relative to a Poisson distribution, the zero-inflated Poisson (ZIP) regression model has been used extensively as a possible machanism for analyzing such data. Böhning [4] reviewed the related literature and provided a variety of biomedical examples.

The ZIP regression model is given by

$$f(y_t|x_t, z_t) = \begin{cases} \omega_t + (1-\omega_t)e^{-\lambda_t}, & \text{if } y_t = 0 \\ (1-\omega_t)e^{-\lambda_t}\lambda_t^k/k!, & \text{if } y_t > 0 \end{cases} \tag{2.1.1}$$

where $0 < \omega_t < 1$,

$$log(\lambda_t) = x_t^T\beta = \textstyle\sum_{j=1}^{k} \beta_j x_{tj},$$

and

$$logit(\omega_t) = log(\omega_t/(1-\omega_t)) = z_t^T\gamma = \textstyle\sum_{j=1}^{m} \gamma_j z_{tj},$$

$x_t^T = (x_{t1}, \cdots, x_{tk})$ is the tth row of covariate matrix X, $\beta = (\beta_1, \cdots, \beta_k)^T$ are unknown k-dimensional column vector of parameters, $z_t^T = (z_{t1}, \cdots, z_{tm})$ is the tth row of covariate matrix Z and $\gamma = (\gamma_1, \cdots, \gamma_m)^T$ are unknown m-dimensional column vector of parameters.

The covariates that affect the probability of the zero state ($Y_t \sim 0$) may or may not be the same as the covariates that affect the Poisson mean of the Poisson state

($Y_t \sim$ Poisson(λ_t)). If the probability of the zero state does not depend on any covariates, then the covariate matrix Z is a column vector of ones, and the ZIP regression requires only one more parameter than Poisson regression. The mean and the variance of the ZIP model [2.1.1] are given, respectively, by

$$EY_t = (1-\omega_t)\lambda_t \quad \text{and} \quad Var(Y_t) = \lambda_t(1-\omega_t)(1+\omega_t\lambda_t)$$

and so this framework also accommodates over-dispersion of the data. This over-dispersion does not arise from heterogeneity, as in the case when the Poisson model is generalized to the Negative Binomial model. Instead, it arises from the splitting of the data into two states. In practice, the presence of over-dispersion may come from one or both of these sources.

2.1.1 Poisson parameter driven model

Poisson regression model is a popular generalized linear model for count data. However, it assumes that the observations are independent. Zeger [67], introduced a regression model for time series of counts assuming that the correlation between observations arises from a latent process added to the linear predictor in log linear model. To proceed, consider a stationary autoregressive process of order p,(AR(p)), such that

$$\alpha_t = \phi_1\alpha_{t-1} + \phi_2\alpha_{t-2} + \cdots + \phi_p\alpha_{t-p} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a normal random process with mean zero and variance σ^2. Conditioning on α_t, suppose Y_t is a sequence of independent counts with Poisson distribution defined as follows:

$$f(y_t|\alpha_t, x_t) = \frac{e^{-\lambda_t}\lambda_t^{y_t}}{y_t!}, \quad y_t = 0, 1, 2, \cdots \tag{2.1.2}$$

and

$$log(\lambda_t) = x_t^T\beta + \alpha_t,$$

where x_t and β are defined as mentioned before in Poisson regression model. Also, assume that

$$f(y_t|\alpha_t) = f(y_t|\alpha_t, \boldsymbol{\alpha}^{(t-1)}) = f(y_t|\alpha_t, \boldsymbol{\alpha}^{(t-1)}, \mathbf{y}^{(t-1)}), \quad t = 1, 2, \cdots \tag{2.1.3}$$

where $\mathbf{y}^{(t)} = (y_t, y_{t-1}, \cdots, y_1)$ and $\boldsymbol{\alpha}^{(t)} = (\alpha_t, \alpha_{t-1}, \cdots, \alpha_0, \alpha_{-1}, \cdots, \alpha_{1-p})$.

Using results in Zeger [67], the marginal moments of the observed process $\{Y_t\}$ are given as follows:

$$EY_t = E(E(Y_t|\alpha_t)) = E(\lambda_t) = e^{x_t^T\beta}Ee^{\alpha_t} = \mu_\alpha\ e^{x_t^T\beta},$$

where $\mu_\alpha = Ee^{\alpha_t}$.

$$\begin{aligned}
\sigma_{y_t}^2 = Var(Y_t) =& E(Var(Y_t|\alpha_t)) + Var(E(Y_t|\alpha_t)) \\
=& E(\lambda_t) + Var(\lambda_t) \\
=& e^{x_t^T\beta}Ee^{\alpha_t} + e^{2x_t^T\beta}Var(e^{\alpha_t}) \\
=& (\mu_\alpha + \sigma_\alpha^2 e^{x_t^T\beta})e^{x_t^T\beta},
\end{aligned}$$

where $\sigma_\alpha^2 = Var(e^{\alpha_t})$.

$$\begin{aligned}
Cov(Y_t, Y_{t+h}) =& E\left(Cov(Y_t|\boldsymbol{\alpha}^{(t+h)}, Y_{t+h}|\boldsymbol{\alpha}^{(t+h)})\right) + Cov(E(Y_t|\boldsymbol{\alpha}^{(t+h)}), E(Y_{t+h}|\boldsymbol{\alpha}^{(t+h)})) \\
=& E\left(Cov(Y_t|\alpha_t, Y_{t+h}|\alpha_{t+h})\right) + Cov(E(Y_t|\alpha_t), E(Y_{t+h}|\alpha_{t+h})) \\
=& Cov(\lambda_t, \lambda_{t+h}) \\
=& e^{(x_t+x_{t+h})^T\beta}Cov(e^{\alpha_t}, e^{\alpha_{t+h}}) \\
=& e^{(x_t+x_{t+h})^T\beta}\ \gamma_\alpha(h),
\end{aligned}$$

where $\gamma_\alpha(h)$ is the autocovariance function of the latent process $\{e^{\alpha_t}\}$. In the above derivation, we have used the fact that $Y_t|\boldsymbol{\alpha}^{(t+h)}$ has the same distribution as $Y_t|\alpha_t$, proven in Proposition B.1. in the Appendix B.

The covariance formula above shows that $\{Y_t\}$ is not a stationary time series since its autocovariance function, $\gamma_{y_t}(h)$, depends on t.

Assuming that $\alpha_t = \phi\alpha_{t-1} + \epsilon_t$ is an AR(1) process, the marginal moments of Y_t are obtained as follows:

$$E(Y_t) = e^{\frac{\sigma^2}{2(1-\phi^2)}+x_t^T\beta},$$

$$\begin{aligned}\sigma^2_{y_t} = Var(Y_t) &= \left[e^{\frac{\sigma^2}{2(1-\phi^2)}} + e^{\frac{\sigma^2}{(1-\phi^2)}+x_t^T\beta}\left(e^{\frac{\sigma^2}{(1-\phi^2)}} - 1\right)\right]e^{x_t^T\beta} \\ &= \left[1 + e^{\frac{\sigma^2}{2(1-\phi^2)}+x_t^T\beta}\left(e^{\frac{\sigma^2}{(1-\phi^2)}} - 1\right)\right]e^{\frac{\sigma^2}{2(1-\phi^2)}+x_t^T\beta},\end{aligned}$$

and

$$\begin{aligned}\gamma_{y_t}(h) = Cov(Y_t, Y_{t+h}) =& e^{(x_t+x_{t+h})^T\beta}\left[e^{\frac{\sigma^2}{(1-\phi^2)}}\left(e^{\frac{\sigma^2\phi^h}{(1-\phi^2)}} - 1\right)\right] \\ =& e^{(x_t+x_{t+h})^T\beta+\frac{\sigma^2}{(1-\phi^2)}}\left(e^{\frac{\sigma^2\phi^h}{(1-\phi^2)}} - 1\right).\end{aligned}$$

Note that, if $\phi > 0$, then there is always positive correlation between Y_t and Y_{t+h}.

2.1.2 ZIP parameter driven model

Although the ZIP regression model has received considerable attention in the literature, it is not suitable for correlated data and especially for time series of count events due to the correlation between successive observations. A zero-

inflated Poisson model with latent process $\{\alpha_t\}$, is introduced to handle such correlation. To be precise we consider a stationary autoregressive process of order p,(AR(p)), such that

$$\alpha_t = \phi_1\alpha_{t-1} + \phi_2\alpha_{t-2} + \cdots + \phi_p\alpha_{t-p} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a normal random process with mean zero and variance σ^2. Conditioning on α_t, suppose Y_t is a sequence of independent counts with ZIP distribution defined as follows:

$$f(y_t|\alpha_t, x_t) = \begin{cases} \omega + (1-\omega)e^{-\lambda_t}, & \text{if } y_t = 0 \\ (1-\omega)e^{-\lambda_t}\lambda_t^{y_t}/y_t!, & \text{if } y_t > 0 \end{cases} \tag{2.1.4}$$

where

$$log(\lambda_t) = x_t^T\beta + \alpha_t,$$

x_t and β are defined as mentioned before in Model [2.1.1], and $0 < \omega < 1$. Also, assume the validity of condition 2.1.3 in Model [2.1.2].

The marginal moments of the observed process $\{Y_t\}$ are given as follows:

$$EY_t = E(E(Y_t|\alpha_t)) = E((1-\omega)\lambda_t) = (1-\omega)e^{x_t^T\beta}Ee^{\alpha_t} = \mu_\alpha\ \mu_t,$$

where $\mu_\alpha = Ee^{\alpha_t}$ and $\mu_t = (1-\omega)e^{x_t^T\beta}$.

$$\begin{aligned}
\sigma_{y_t}^2 = Var(Y_t) =& E(Var(Y_t|\alpha_t)) + Var(E(Y_t|\alpha_t)) \\
=& E(\lambda_t(1-\omega)(1+\lambda_t\omega)) + Var((1-\omega)\lambda_t) \\
=& (1-\omega)\{e^{x_t^T\beta}Ee^{\alpha_t} + \omega e^{2x_t^T\beta}Ee^{2\alpha_t}\} + (1-\omega)^2e^{2x_t^T\beta}Var(e^{\alpha_t}) \\
=& (1-\omega)\{\mu_\alpha e^{x_t^T\beta} + \omega e^{2x_t^T\beta}(Var(e^{\alpha_t}) + \mu_\alpha^2)\} + (1-\omega)^2e^{2x_t^T\beta}Var(e^{\alpha_t}) \\
=& (1-\omega)\{\mu_\alpha e^{x_t^T\beta} + \omega e^{2x_t^T\beta}\mu_\alpha^2 + e^{2x_t^T\beta}Var(e^{\alpha_t})\}
\end{aligned}$$

$$=\mu_\alpha\ \mu_t + \frac{\omega}{1-\omega}\ \mu_\alpha^2\ \mu_t^2 + \frac{1}{1-\omega}\mu_t^2\sigma_\alpha^2,$$

where $\sigma_\alpha^2 = Var(e^{\alpha_t})$.

$$\begin{aligned}
\gamma_{y_t}(h) =& E\left(Cov(Y_t|\boldsymbol{\alpha}^{(t+h)}, Y_{t+h}|\boldsymbol{\alpha}^{(t+h)})\right) + Cov(E(Y_t|\boldsymbol{\alpha}^{(t+h)}), E(Y_{t+h}|\boldsymbol{\alpha}^{(t+h)})) \\
=& E\left(Cov(Y_t|\alpha_t, Y_{t+h}|\alpha_{t+h})\right) + Cov(E(Y_t|\alpha_t), E(Y_{t+h}|\alpha_{t+h})) \\
=& Cov((1-\omega)\lambda_t, (1-\omega)\lambda_{t+h}) \\
=& (1-\omega)^2\ e^{x_t^T\beta} e^{x_{t+h}^T\beta} Cov(e^{\alpha_t}, e^{\alpha_{t+h}}) \\
=& \mu_t\ \mu_{t+h}\ \gamma_\alpha(h),
\end{aligned}$$

where $\gamma_\alpha(h)$ is the autocovariance function of the latent process $\{e^{\alpha_t}\}$. Note that $\{Y_t\}$ is not a stationary time series as to be expected, since $\gamma_{y_t}(h)$ is not free of t.

At this point, a few remarks are in order. Firstly, we notice that the ZIP parameter-driven model defined here can also accommodate overdispersion because the mean function is always smaller than the variance function. Secondly, both negative and positive autocorrelations are accommodated depending on the parameters of $\gamma_\alpha(h)$, the autocovariance function of the latent process.

As an example, if the latent process is a Gaussian $AR(1)$, $\alpha_t = \phi\alpha_{t-1} + \epsilon_t$, then the marginal moments of Y_t reduce to

$$E(Y_t) = (1-\omega)e^{\frac{\sigma^2}{2(1-\phi^2)} + x_t^T\beta},$$

$$\begin{aligned}
\sigma_{y_t}^2 = Var(Y_t) =& (1-\omega)e^{\frac{\sigma^2}{2(1-\phi^2)} + x_t^T\beta}\left[1 + \omega e^{\frac{\sigma^2}{2(1-\phi^2)} + x_t^T\beta} + e^{\frac{\sigma^2}{2(1-\phi^2)} + x_t^T\beta}(e^{\frac{\sigma^2}{1-\phi^2}} - 1)\right] \\
=& (1-\omega)e^{\frac{\sigma^2}{2(1-\phi^2)} + x_t^T\beta}\left[1 + e^{\frac{\sigma^2}{2(1-\phi^2)} + x_t^T\beta}(\omega + e^{\frac{\sigma^2}{1-\phi^2}} - 1)\right],
\end{aligned}$$

and

$$\gamma_{y_t}(h) = (1-\omega)^2 e^{(x_t+x_{t+h})^T\beta+\frac{\sigma^2}{1-\phi^2}}(e^{\frac{\sigma^2\phi^h}{1-\phi^2}}-1).$$

Notice that the autocovariance function could be negative or positive according to the values of ϕ.

2.2 Estimation

Let $y = (y_1, \cdots, y_n)^T$ be the observed data vector, and conditionally on the latent autoregressive process $\{\alpha_t\}$, we assume that the elements of y are independent with parameters $\theta = (\theta_1, \theta_2)$, where $\theta_1 = \beta$ denotes the parameters of the fixed effects when y is assumed to be drawn from model [2.1.2], and $\theta_1 = (\beta, \omega)$ denotes the parameters of the fixed effects and the zero state probability when y is assumed to be drawn from model [2.1.4]. $\theta_2 = (\phi_1, \cdots, \phi_p, \sigma)$ denotes the parameters of the autoregressive latent process in both models.

2.2.1 Parameter estimation of Poisson model

We need to write the Poisson model [2.1.2] as a hierarchical model in order to estimate the parameters of the posterior distribution using DC method. To proceed, consider the following model

$$Y_t|x_t, \alpha_t, \sim \text{ Poisson}(\lambda_t), \text{ with } \lambda_t = \exp(x_t^T\beta + \alpha_t),$$

$$\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p} \sim \text{Normal}(\phi_1\alpha_{t-1} + \cdots, \phi_p\alpha_{t-p}, \sigma^2).$$

The likelihood function of this model is obtained by

$$L(\theta, y) = \int \prod_{t=1}^{n} f(y_t|\alpha_t)g(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p})g_0(\alpha_0, \cdots, \alpha_{1-p})\, d\alpha,$$

where $\alpha_0, \cdots, \alpha_{1-p}$ are the initial conditions of the process $\{\alpha_t\}$. The derivation of this likelihood function is outlined in Appendix A.

The posterior distribution of θ conditional on the data $y^{(K)} = (y, \cdots, y)$ is given by

$$\pi_K(\theta|y) = \frac{[L(\theta, y)]^K \pi(\theta)}{C(K, y)},$$

where $C(K, y) = \int [L(\theta, y)]^K \pi(\theta) d\theta$, is the normalizing constant.

2.2.2 Parameter estimation of the ZIP model

To apply the DC algorithm, we need to write the ZIP model [2.1.4] as a hierarchical model. Let u_t be a random variable such that $u_t = 0$ when Y_t is from the zero state, and $u_t = 1$ if Y_t is from the Poisson state. Then the ZIP model can be written as follows:

$$Y_t|x_t, \alpha_t, u_t \sim \text{Poisson}(u_t\lambda_t + 0.000001), \text{ with } \lambda_t = \exp(x_t^T\beta + \alpha_t),$$

$$u_t \sim \text{Bernoulli}(1 - \omega),$$

$$\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p} \sim \text{Normal}(\phi_1\alpha_{t-1} + \cdots + \phi_p\alpha_{t-p}, \sigma^2).$$

Note that if $u_t = 0$, then

$$P(Y_t = 0|\alpha_t, u_t = 0) = 1 \quad \text{and} \quad P(Y_t = k|\alpha_t, u_t = 0) = 0, \quad k > 0,$$

and if $u_t = 1$, then

$$P(Y_t = k|\alpha_t, u_t = 1) = e^{-\lambda_t}\lambda_t^k/k!, \quad k = 0, 1, 2, \cdots$$

Hence,

$$P(Y_t = 0|\alpha_t) = P(u_t = 0)P(Y_t = 0|\alpha_t, u_t = 0) + P(u_t = 1)P(Y_t = 0|\alpha_t, u_t = 1)$$

$$=\omega + (1-\omega)e^{-\lambda_t},$$

and for $k > 0$,

$$\begin{aligned} P(Y_t = k|\alpha_t) =& P(u_t = 0)P(Y_t = k|\alpha_t, u_t = 0) + P(u_t = 1)P(Y_t = k|\alpha_t, u_t = 1) \\ =& (1-\omega)e^{-\lambda_t}\lambda_t^k/k!, \end{aligned}$$

which shows that Y_t has the ZIP distribution [2.1.4]. Of course, adding the number 0.000001 to the Poisson mean is just a trick to avoid the case that the Poisson mean is exactly zero.

The likelihood function of this model is computed as follows:

$$L(\theta, y, \boldsymbol{\alpha}^{(n)}) = \prod_{t=1}^{n} f(y_t|\alpha_t)g(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p})g_0(\alpha_0, \cdots, \alpha_{1-p}),$$

for details see Appendix A. Also, $f(y_t|\alpha_t) = \sum_{u_t=0}^{1} h(u_t)f(y_t|\alpha_t, u_t)$. Hence, the likelihood function of the model is

$$L(\theta, y) = \int \prod_{t=1}^{n} \left[\sum_{u_t=0}^{1} h(u_t)f(y_t|\alpha_t, u_t)\right] g(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p})g_0(\alpha_0, \cdots, \alpha_{1-p})\ d\alpha.$$

The posterior distribution of θ conditional on the data $y^{(K)} = (y, \cdots, y)$ is given by

$$\pi_K(\theta|y) = \frac{[L(\theta, y)]^K \pi(\theta)}{C(K, y)},$$

where $C(K, y) = \int [L(\theta, y)]^K \pi(\theta)d\theta$, is the normalizing constant.

2.3 Forecasting

For ZIP parameter-driven model 2.1.4, and following Section 8.8 of Brockwell and Davis [7], the forecast density of the next observation Y_{n+1} given the current data $Y^{(n)}$ can be computed recursively. To illustrate

$$\begin{aligned} E(Y_{n+1}|Y^{(n)}) =& E\left[E(Y_{n+1}|\alpha_{n+1})|Y^{(n)}\right] \\ =& E\left[(1-\omega)e^{x_{n+1}^T\beta+\alpha_{n+1}}|Y^{(n)}\right] \\ =& E\left[E((1-\omega)e^{x_{n+1}^T\beta+\alpha_{n+1}}|\alpha_n,\cdots,\alpha_{n-p+1})|Y^{(n)}\right] \\ =& E\left[(1-\omega)e^{x_{n+1}^T\beta+\phi_1\alpha_n+\cdots+\phi_p\alpha_{n-p+1}+\frac{1}{2}\sigma^2}|Y^{(n)}\right]. \end{aligned}$$

To compute the last conditional expectation, it is enough to generate a large number of replicates $\boldsymbol{\alpha}_i = (\alpha_n,\cdots,\alpha_{n-p+1})^i, i = 1,\cdots,N$, computed from the conditional distribution of $\boldsymbol{\alpha}^{(n)}$ given $\mathbf{Y}^{(n)}$ such that

$$p(\boldsymbol{\alpha}^{(n)}|\mathbf{y}^{(n)}) = \frac{\prod_{i=1}^n p(y_i|\alpha_i)p(\alpha_i|\alpha_{i-1},\cdots,\alpha_{i-p})p(\alpha_0,\cdots,\alpha_{1-p})}{\int\prod_{i=1}^n p(y_i|\alpha_i)p(\alpha_i|\alpha_{i-1},\cdots,\alpha_{i-p})p(\alpha_0,\cdots,\alpha_{1-p})d\alpha},$$

and then approximate the conditional expectation by

$$E(Y_{n+1}|Y^{(n)}) = \frac{\sum_{i=1}^N(1-\omega)e^{x_{n+1}^T\beta+\boldsymbol{\alpha}_i^T\phi+\frac{1}{2}\sigma^2}}{N},$$

where $\phi = (\phi_1,\cdots,\phi_p)$. Of course, the same ideas can be applied for predicting the counts times further into the future.

In general, one can compute $E(h(\alpha_n,\cdots,\alpha_{n-p+1})|Y^{(n)})$, where $h(\alpha_n,\cdots,\alpha_{n-p+1})$ is a function of $\alpha_n,\cdots,\alpha_{n-p+1}$, by generating a large number of replicates computed from the conditional distribution $\boldsymbol{\alpha}^{(n)}$ given $\mathbf{Y}^{(n)}$ and then approximate the

conditional expectation empirically,

$$E(h(\alpha_n, \cdots, \alpha_{n-p+1})|Y^{(n)}) = \frac{\sum_{i=1}^{N} h(\boldsymbol{\alpha}_i)}{N}.$$

The following proposition derives the conditional variance, $var(Y_{n+1}|Y^{(n)})$.

Proposition 2.3.1.

$$var(Y_{n+1}|Y^{(n)}) = (1-\omega)^2 e^{2x_{n+1}^T\beta}\left[\frac{\sum_{i=1}^{N} e^{2(\boldsymbol{\alpha}_i^T\phi+\sigma^2)}}{N} - \left(\frac{\sum_{i=1}^{N} e^{\boldsymbol{\alpha}_i^T\phi+\frac{1}{2}\sigma^2}}{N}\right)^2\right]$$

$$+(1-\omega)\left[\frac{\sum_{i=1}^{N}(e^{x_{n+1}^T\beta+\boldsymbol{\alpha}_i^T\phi+\frac{1}{2}\sigma^2)} + e^{2(x_{n+1}^T\beta+\boldsymbol{\alpha}_i^T\phi+\sigma^2)})}{N}\right],$$

Proof. The conditional variance of the next observation Y_{n+1} given the current data $Y^{(n)}$ in Section 2.3 is computed as follows:

$$var(Y_{n+1}|Y^{(n)}) = E\left[var(Y_{n+1}|\alpha_{n+1})|Y^{(n)}\right] + var\left[E(Y_{n+1}|\alpha_{n+1})|Y^{(n)}\right]$$

$$= E\left[(1-\omega)e^{x_{n+1}^T\beta+\alpha_{n+1}}(1+\omega e^{x_{n+1}^T\beta+\alpha_{n+1}})|Y^{(n)}\right] + var\left[(1-\omega)e^{x_{n+1}^T\beta+\alpha_{n+1}}|Y^{(n)}\right]$$

$$= E\left[E(1-\omega)e^{x_{n+1}^T\beta+\alpha_{n+1}}(1+\omega e^{x_{n+1}^T\beta+\alpha_{n+1}})|\alpha_n, \cdots, \alpha_{n-p+1})|Y^{(n)}\right] +$$

$$(1-\omega)^2 e^{2x_{n+1}^T\beta}\left[var\left(E(e^{\alpha_{n+1}}|\alpha_n, \cdots, \alpha_{n-p+1})|Y^{(n)}\right) + E\left(var(e^{\alpha_{n+1}}|\alpha_n, \cdots, \alpha_{n-p+1})|Y^{(n)}\right)\right].$$

Note that,

$$\alpha_{n+1}|\alpha_n, \cdots, \alpha_{n-p+1} \sim \text{Normal}(\phi_1\alpha_n + \cdots + \phi_p\alpha_{n-p+1}, \sigma^2),$$

hence,

$E(e^{\alpha_{n+1}}|\alpha_n, \cdots, \alpha_{n-p+1}) = e^{\phi_1\alpha_n+\cdots+\phi_p\alpha_{n-p+1}+\frac{1}{2}\sigma^2}$,

$E(e^{2\alpha_{n+1}}|\alpha_n, \cdots, \alpha_{n-p+1}) = e^{2(\phi_1\alpha_n+\cdots+\phi_p\alpha_{n-p+1}+\sigma^2)}$,

and

$var(e^{\alpha_{n+1}}|\alpha_n, \cdots, \alpha_{n-p+1}) = e^{2(\phi_1\alpha_n+\cdots+\phi_p\alpha_{n-p+1})+\sigma^2}(e^{\sigma^2}-1)$.

Consequently, the conditional variance becomes

$$var(Y_{n+1}|Y^{(n)}) = (1-\omega)^2 e^{2x_{n+1}^T\beta}[var(e^{\boldsymbol{\alpha}^T\phi+\frac{1}{2}\sigma^2})|Y^{(n)})+E(e^{2(\boldsymbol{\alpha}^T\phi+\sigma^2)}-e^{2\boldsymbol{\alpha}^T\phi+\sigma^2}|Y^{(n)})]$$

$$+E[(1-\omega)(e^{x_{n+1}^T\beta+\boldsymbol{\alpha}^T\phi+\frac{1}{2}\sigma^2}+e^{2(x_{n+1}^T\beta+\boldsymbol{\alpha}^T\phi+\sigma^2)})|Y^{(n)}]$$

which can be written as

$$var(Y_{n+1}|Y^{(n)}) = (1-\omega)^2 e^{2x_{n+1}^T\beta}[E(e^{2(\boldsymbol{\alpha}^T\phi+\sigma^2)}|Y^{(n)}) - [E(e^{\boldsymbol{\alpha}^T\phi+\frac{1}{2}\sigma^2}|Y^{(n)})]^2$$

$$+(1-\omega)E[e^{x_{n+1}^T\beta+\boldsymbol{\alpha}^T\phi+\frac{1}{2}\sigma^2}+e^{2(x_{n+1}^T\beta+\boldsymbol{\alpha}^T\phi+\sigma^2)}|Y^{(n)}],$$

where $\boldsymbol{\alpha} = (\alpha_n, \cdots, \alpha_{n-p+1})$ and $\phi = (\phi_1, \cdots, \phi_p)$.

□

2.4 Numerical studies

We considered two experiments: one with a Poisson and the other with a ZIP for the conditional distribution of observations given the latent process. For each case, we simulated 500 realizations and estimated the parameters of interest, reporting the empirical means, the empirical standard deviations and the mean square errors (MSE) of the estimates together with the asymptotic standard deviation.

2.4.1 Experiment 1: Poisson.

Tables 2.1 and 2.3 show the real values of the parameters of model [2.1.2] from which the data has been simulated, the empirical means, the empirical standard deviations, MSE and DC standard errors with ($K = 5$). With the AR(1) latent

process we used a covariate sequence defined by $x_t^T = (1, sin\frac{\pi t}{2}, cos\frac{\pi t}{2})$, which includes two harmonic functions components, and with the AR(2) latent process we used $x_t^T = (1, \frac{t}{n})$, which includes standardized trend. In both cases we set the sample size to be 500.

Tables 2.2 and 2.4 show the empirical means, the empirical standard deviations, MSE and Bayes MCMC standard errors with $(K = 1)$.

We used the following priors in our simulations: normal distribution with mean zero and variance 10^4 for fixed effects parameters, log normal distribution with mean 0 and variance 1 for the inverse of the variance component. For the latent process AR(1), we used uniform prior distribution on the interval $(-0.99, 0.99)$ for parameter ϕ_1, and for the parameters ϕ_1 and ϕ_2 in the latent process AR(2) we used normal prior distribution with mean 0 and variance 10^4. Also, we assigned normal prior distributions with mean zero and variance one for the initial condition parameters in both processes.

In each case we set the following: a burn-in period of 2000, two parallel MCMC chains and 5000 values to generate from the posterior distribution from each chain.

Table 2.5 gives the percentage of coverage when $\alpha = 0.05$, that is the proportion of times the $100(1 - \alpha)\%$ confidence interval $\widehat{\theta_i} \pm z_{\frac{\alpha}{2}} SE(\widehat{\theta_i})$ contains the true parameter θ_i. The table shows similar accurate results with AR(1) and AR(2) latent processes.

Table 2.1: Estimation of Poisson model parameters with AR(1) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_0	-0.50	-0.4930	0.0112	0.1057	0.1117
β_1	1.00	0.9986	0.0090	0.0948	0.0975
β_2	0.70	0.6959	0.0090	0.0950	0.0958
σ	1.00	0.9924	0.0055	0.0741	0.0753
ϕ_1	0.45	0.4390	0.0057	0.0749	0.0780

Table 2.2: Estimation of Poisson model parameters with AR(1) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_0	-0.50	-0.5010	0.0113	0.0865	0.1136
β_1	1.00	1.0020	0.0090	0.0747	0.0983
β_2	0.70	0.6984	0.0091	0.0760	0.0965
σ	1.00	1.0020	0.0055	0.0587	0.0760
ϕ_1	0.45	0.4355	0.0058	0.0597	0.0782

Table 2.3: Estimation of Poisson model parameters with AR(2) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_0	0.50	0.5031	0.0264	0.1626	0.1675
β_1	2.00	1.9981	0.0697	0.2643	0.2727
σ	0.80	0.7945	0.0016	0.0400	0.0416
ϕ_1	1.00	0.9985	0.0026	0.0514	0.0539
ϕ_2	-0.50	-0.5023	0.0024	0.0485	0.0516

Table 2.4: Estimation of Poisson model parameters with AR(2) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_0	0.50	0.5028	0.0281	0.1676	0.1614
β_1	2.00	1.9947	0.0768	0.2773	0.2678
σ	0.80	0.7998	0.0016	0.0404	0.0415
ϕ_1	1.00	0.9972	0.0026	0.0512	0.0540
ϕ_2	-0.50	-0.4983	0.0024	0.0489	0.0512

Table 2.5: Percentage of coverage of a 95% confidence interval in Poisson model

Parameters	AR(1) latent process	Parameters	AR(2) latent process
β_0	0.960	β_0	0.930
β_1	0.946	β_1	0.934
β_2	0.946	σ	0.954
σ	0.940	ϕ_1	0.960
ϕ_1	0.956	ϕ_2	0.956

2.4.2 Experiment 2: ZIP.

Tables 2.6 and 2.8 show the real values of the parameters of model [2.1.4] from where the data have been simulated, the empirical means, the empirical standard deviations, MSE and DC standard errors with ($K = 3$).

With the latent process AR(1), we used a covariate sequence $x_t^T = (1, x_{t2})$, where x_{t2} is standard normal random variable, and with the latent process AR(2), we used a sequence of normal random variables with mean 0.5 and variance 0.25. A sample size of 1000 was used in both cases.

Tables 2.7 and 2.9 show the empirical means, the empirical standard deviations, MSE and Bayes MCMC standard errors with ($K = 1$).

The priors we used for the parameters β, ω and σ were $uniform(-2.5, 2.5)$, $Beta(1, 1)$ and $log\ Normal(0, 1)$. For ϕ_1 in the AR(1) latent process we used $uniform(-0.99, 0.99)$ and $normal(0, 10^4)$ for ϕ_1 and ϕ_2 in the AR(2) process. Also, we set the normal prior distribution with mean zero and variance 1 for the ini-

tial condition parameters in both processes. The burn-in period was 2000, three parallel MCMC chains were generated with 10000 iterations in each one.

Tables 2.6 and 2.7 show almost similar accurate results except for the parameter ϕ_1, we see that DC method gives an estimator with percentage bias of 0.05 which is almost half the percentage bias obtained by Bayes MCMC which is 0.09. There is also good agreement between the standard deviations observed over the replications and the standard errors obtained by both DC and Bayesian MCMC.

Furthermore,The percentage of coverage of the MLE using DC method is shown in Table 2.10.

Table 2.6: Estimation of ZIP model parameters with AR(1) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_0	1.00	1.0031	0.0019	0.0437	0.0472
β_1	0.50	0.4986	0.0011	0.0324	0.0320
ω	0.35	0.3491	0.0004	0.0193	0.0194
σ	0.40	0.3892	0.0024	0.0475	0.0569
ϕ_1	0.40	0.3791	0.0161	0.1253	0.1476

Table 2.7: Estimation of ZIP model parameters with AR(1) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_0	1.00	0.9983	0.0019	0.0434	0.0446
β_1	0.50	0.4991	0.0011	0.0325	0.0320
ω	0.35	0.3484	0.0003	0.0194	0.0190
σ	0.40	0.3953	0.0022	0.0468	0.0468
ϕ_1	0.40	0.3622	0.0160	0.1207	0.1344

Table 2.8: Estimation of ZIP model parameters with AR(2) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_0	0.50	0.5005	0.0039	0.0623	0.0607
β_1	1.00	1.0017	0.0049	0.0702	0.0602
ω	0.20	0.2006	0.0003	0.0173	0.0184
σ	0.50	0.4918	0.0020	0.0434	0.0401
ϕ_1	-0.40	-0.4346	0.0133	0.1101	0.1002
ϕ_2	0.45	0.4099	0.0130	0.1072	0.0991

Table 2.9: Estimation of ZIP model parameters with AR(2) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_0	0.50	0.4976	0.0037	0.0607	0.0564
β_1	1.00	1.0026	0.0050	0.0705	0.0595
ω	0.20	0.2005	0.0003	0.0170	0.0176
σ	0.50	0.4945	0.0016	0.0393	0.0361
ϕ_1	-0.40	-0.4408	0.0124	0.1037	0.0963
ϕ_2	0.45	0.4029	0.0124	0.1012	0.0956

Table 2.10: Percentage of coverage of a 95% confidence interval in ZIP model

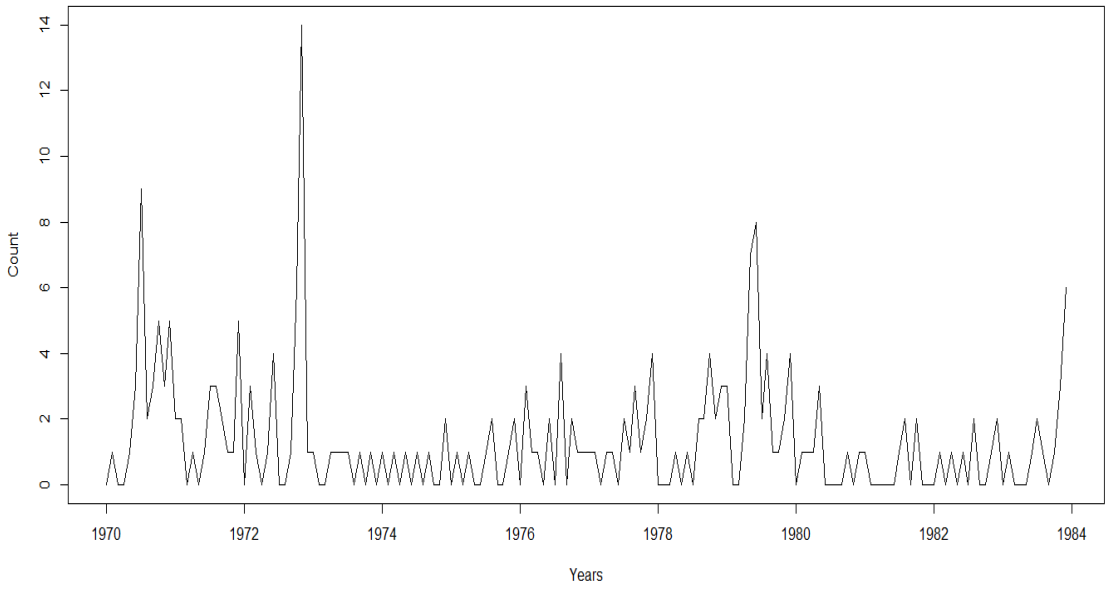
Parameters	AR(1) latent process	AR(2) latent process
β_0	0.966	0.938
β_1	0.950	0.938
ω	0.942	0.958
σ	0.958	0.944
ϕ_1	0.968	0.934
ϕ_2	-	0.940

2.5 Application to real data

2.5.1 Polio dataset

We applied DC method to the polio data example from Zeger [67]. The data consist of monthly counts of poliomyelities cases in the USA from the year 1970 to 1983 as reported by the Center for Disease Control. The data, which is graphed in Figure 2.1 reveals some seasonality and the possibility of a slight decreasing trend. The main objective in modelling this data is the detection of a decreasing trend.

Figure 2.1: Monthly counts of poliomyelities cases in USA. (1970-1983)



We used the same regression variables as in Zeger [67], namely

$$x_t^T = (1, \tilde{t}/1000, cos(2\pi\tilde{t}/12), sin(2\pi\tilde{t}/12), cos(2\pi\tilde{t}/6), sin(2\pi\tilde{t}/6)),$$

where $\tilde{t} = t - 73$ is used to locate the intercept term of January 1976. Also,

we assumed that there is an unobserved Gaussian AR(1) process $\{\alpha_t\}$ satisfying $\alpha_t = \phi_1\alpha_{t-1} + \epsilon_t$, where the ϵ_t are independent and $normal(0, \sigma^2)$. Given the latent process $\{\alpha_t\}$, the observations $y_t, (t = 1, \cdots, 168)$ are independent and following Poisson distribution with mean $\lambda_t = exp(x_t^T\beta + \alpha_t)$.

Table 2.11 reports the model parameter estimates and the corresponding standard errors using DC method with 120 colons, 3 MCMC chains, 5000 burn-in period and 30000 iterations. The model suggests no evidence of decreasing trend in the rate of US polio infections over time. Both annual and semi-annual seasonal effects are statistically significant at 0.001 level. The variance and correlation components are also significant.

Table 2.11: Estimates and their standard errors from analysis of polio data using DC method

Parameter	MLE	DC SE	z-value
β_0	-0.0308	0.1543	-0.20
β_1	-3.8600	2.8522	-1.35
β_2	-0.0974	0.1491	-0.65
β_3	-0.4954	0.1583	-3.13**
β_4	0.1986	0.1249	1.59
β_5	-0.3627	0.1256	-2.89**
ϕ_1	0.6561	0.1646	3.99**
σ	0.5249	0.1235	4.25**

** indicates significant at 0.01 level

To check the convergence of the DC approach, we calculated the largest eigen-

value of the posterior variance-covariance matrix (lambda.max), mean squared error (ms.error) and correlation-like fit statistic (r^2). The maximum eigenvalue reflects the degeneracy of the posterior distribution, while the other two measures reflect the adequacy of the normal approximation. All these statistics should converge to zero as K increases. If this happens, different prior specifications are no longer influencing the results (Lele et al. [42] and [43]). These are conveniently calculated by the function *dcdiag* in *dclone* package in R, Table 2.12 shows these measures. To also further investigate the behaviour of MCMC chains, we computed Brook-Gelman statistic, $\widehat{R}$, and it converges to 1 as shown in Table 2.12, Figure 2.2 shows the posterior distribution for the parameters with (K=110) and it looks appropriately normal, while Figure 2.3 presents trace plots of the model parameters, which suggested that the samples of all the parameters got mixed well.

Table 2.12: Estimability diagnostics

Number of clones	lambda.max	ms error	r^2	$\widehat{R}$
1	12.5213	1.4644	0.0342	1.0061
5	1.7503	0.0370	0.0009	1.0030
10	0.8556	0.0090	0.0002	1.0063
20	0.4169	0.0023	0.0001	1.0030
50	0.1683	0.0016	0.0005	1.0021
80	0.1021	0.0021	0.0001	1.0018
110	0.0740	0.0004	0.0000	1.0055

Figure 2.2: The posterior densities of Poisson model parameters

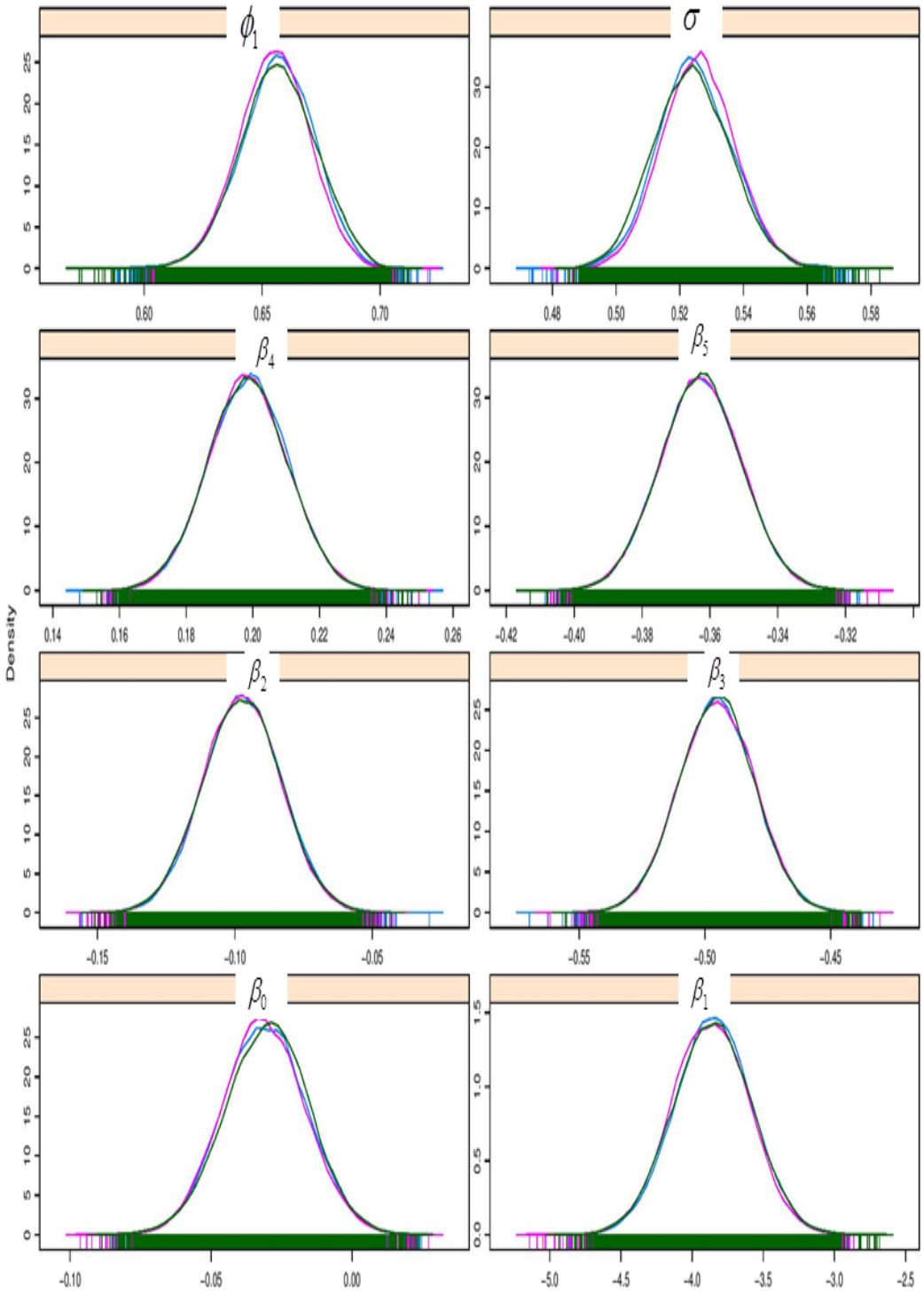
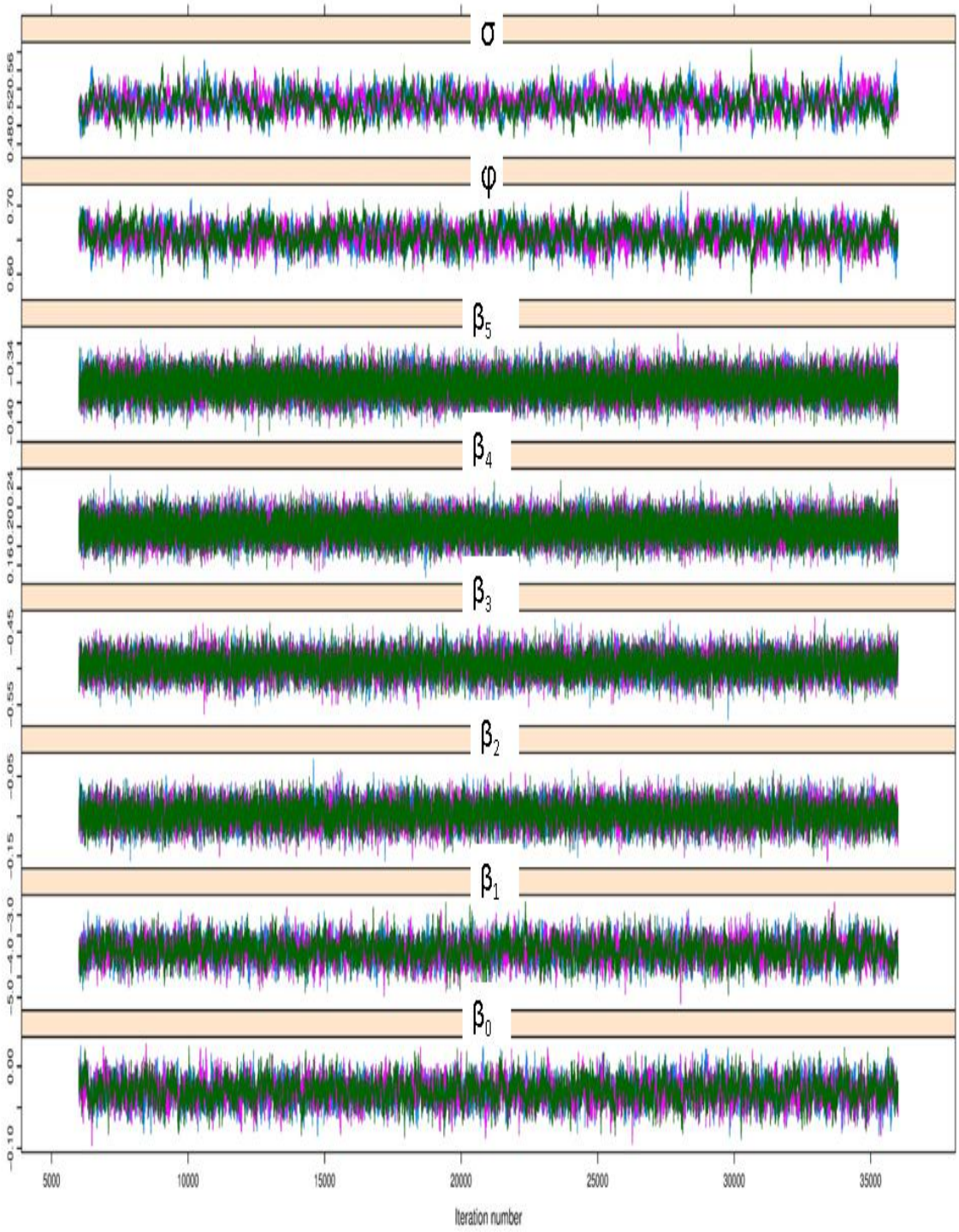


Figure 2.3: The trace plots of Poisson model parameters



2.5.2 Asthma age group (70-79) years dataset

Analysis of the trend of Emergencey Department (ED) visits related to asthma is relevant to public health planning. For illustration purposes, in this section, we use a dataset of daily counts of emergency department visits due to asthma for people in the age group (70-79) years in Ontario, Canada during the period January 1st, 2010 to December 31st, 2015 (sample size = 2191). The data are part of the National Ambulatory Care Reporting System (NACRS) maintained by the Canadian Institute for Health Information (CIHI). Due to CIHI's privacy and confidentiality policies, any counts that are fewer than five visits per day was suppressed. For the purposes of the present application, we considered any count less than 5 to be zero and subtracted 4 from counts greater or equal to 5. Figure 2.4 shows the daily number of asthma visits from January, 1, 2010 till December, 31, 2010 for the age group (70-79) years. Also, Figure 2.5 presents a histogram of these data. These figures clearly show the need of a zero-inflation model.

Figure 2.4: Asthma presentation for age group (70-79) years.

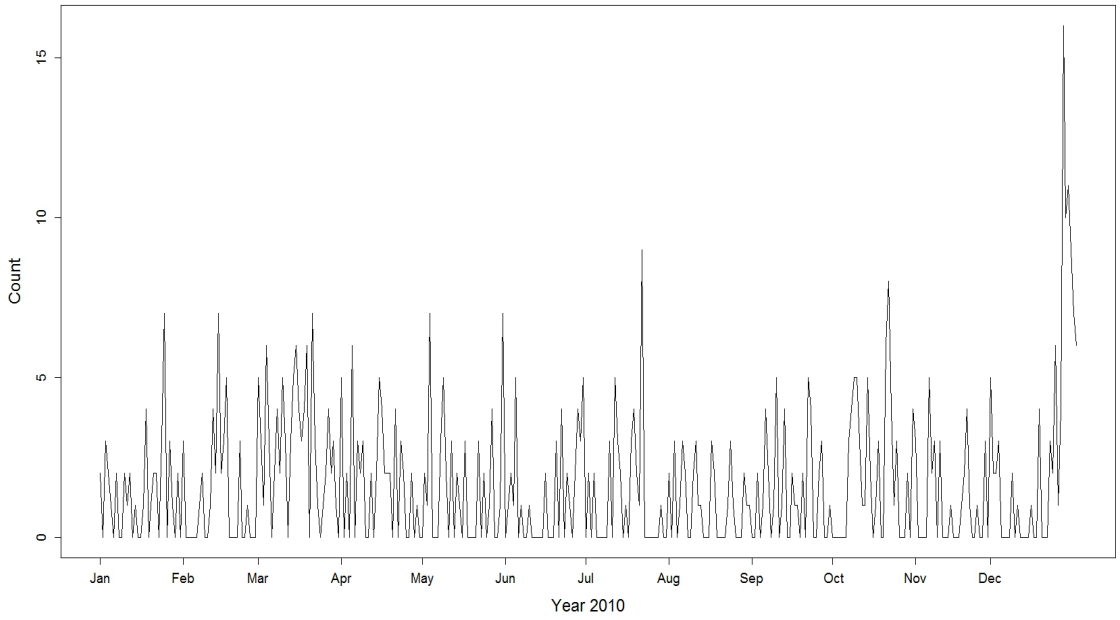
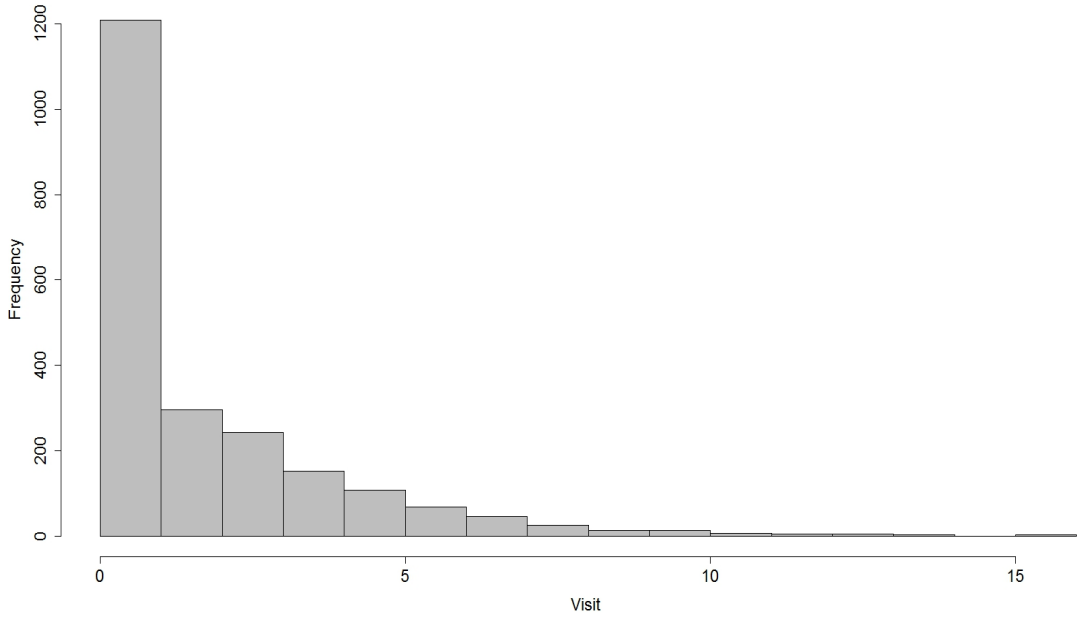


Figure 2.5: Histogram of daily visits of asthma cases of people aged (70-79) years in Ontario. (January 2010 - December 2015)



The data were fitted using the following set of explanatory variables: $x_{1t} = \frac{t}{2191}$, to include time trend; x_{2t} to include weekend effect; $x_{3t} = cos\frac{2\pi t}{365}$; $x_{4t} = sin\frac{2\pi t}{365}$; $x_{5t} = cos\frac{4\pi t}{365}$ and $x_{6t} = sin\frac{4\pi t}{365}$ to include annual and semi-annual pattern. For weekend variable we coded it 1 for Saturday and Sunday and zero elsewhere. The dependent variable is emergency department daily asthma visits by people aged (70-79) years.

We fitted ZIP parameter driven model to asthma data assuming that there is AR(1) latent process, $\alpha_t = \phi_1\alpha_{t-1} + \epsilon_t$, where $\epsilon_t \sim Normal(0, \sigma^2)$. We also fitted ZIP model without any latent process and presented the results in Table 2.13.

Table 2.13: Estimates and their standard errors from analysis of asthma data by ZIP model with and without latent process

Parameter	ZIP parameter driven model			ZIP model		
	MLE	DC SE	z-value	MLE	DC SE	z-value
β_0	0.6238	0.0919	6.79**	0.9411	0.0375	25.10**
β_1	0.2445	0.1451	1.69	0.2460	0.0585	4.21**
β_2	-0.0224	0.0446	-0.50	-0.0329	0.0365	-0.90
β_3	0.2015	0.0606	3.33**	0.2525	0.0259	9.75**
β_4	0.3278	0.0613	5.35**	0.2356	0.0256	9.20**
β_5	-0.0681	0.0595	-1.14	0.0350	0.0235	1.49
β_6	-0.0477	0.0602	-0.79	-0.0927	0.0260	-3.57**
ω	0.2784	0.0161	17.29**	0.3651	0.0116	31.47**
ϕ_1	0.7906	0.0364	21.72**	-	-	-
σ	0.3520	0.0368	9.56**	-	-	-

** indicates significant at 0.01 level

The models were fitted in *jags* and called into R using package *rjags*. We ran the sampler for 50000 iterations, discarding the first 20000 as burn-in and using three MCMC chains. Furthermore, the values of the maximum eigenvalue of the posterior variance-covariance matrix, mean square error, correlation-like fit statistic r^2 and Brooks-Gelman statistic, $\widehat{R}$, with K=3 are 0.0091, 0.0006, 0.0000 and 1.022885, respectively, for ZIP parameter driven model, and 0.0015, 0.0050, 0.0002 and 1.0009, respectively, for ZIP model. These values reflect the degenerateness of the posterior distribution, the adequacy of the normal approximation and the

convergence of the MCMC chains. Figure 2.6 shows the posterior densities of ZIP parameter driven model which look appropriately normal, while Figure 2.7 presents trace plots of ZIP parameter driven model, which suggested that the samples of all the parameters got mixed well.

From Table 2.13 we see that in both models annual seasonal effects are statistically significant at 0.001 level, while weekend effect and semi-annual effects are not significant. In parameter driven model, our estimate $\widehat{\beta}_1 = 0.2445$ with asymptotic standard error 0.1451 is clearly not significant, while in the model without latent process is significant at 0.001 level. The variance component in ZIP parameter driven model is statistically significant at 0.001 level, which means the number of asthma visitors to the emergency department of hospitals for the age group (70-79) years express significant heterogeneity. also, significant correlation is detected at 0.001 level.

Figure 2.6: The posterior densities of ZIP model parameters.

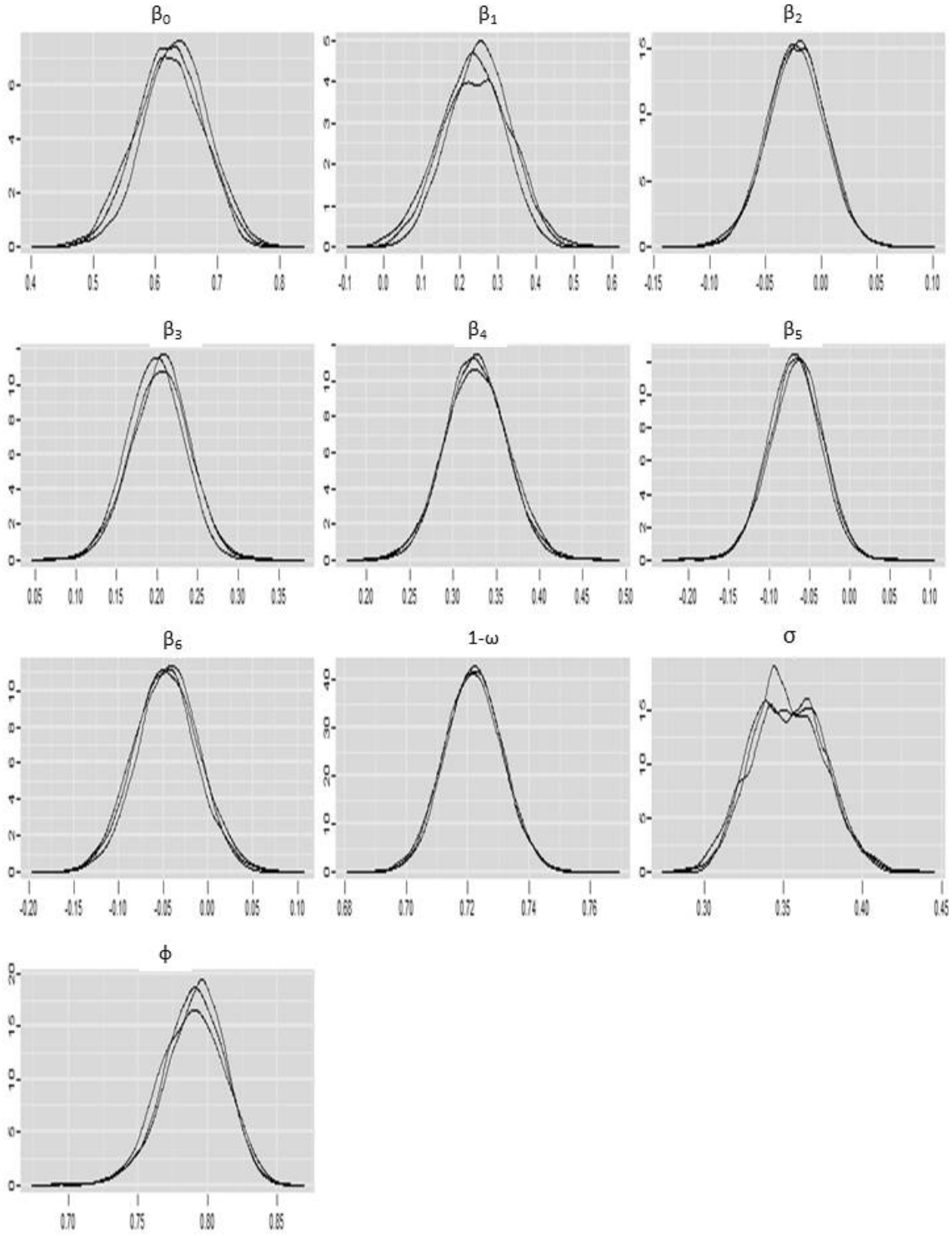
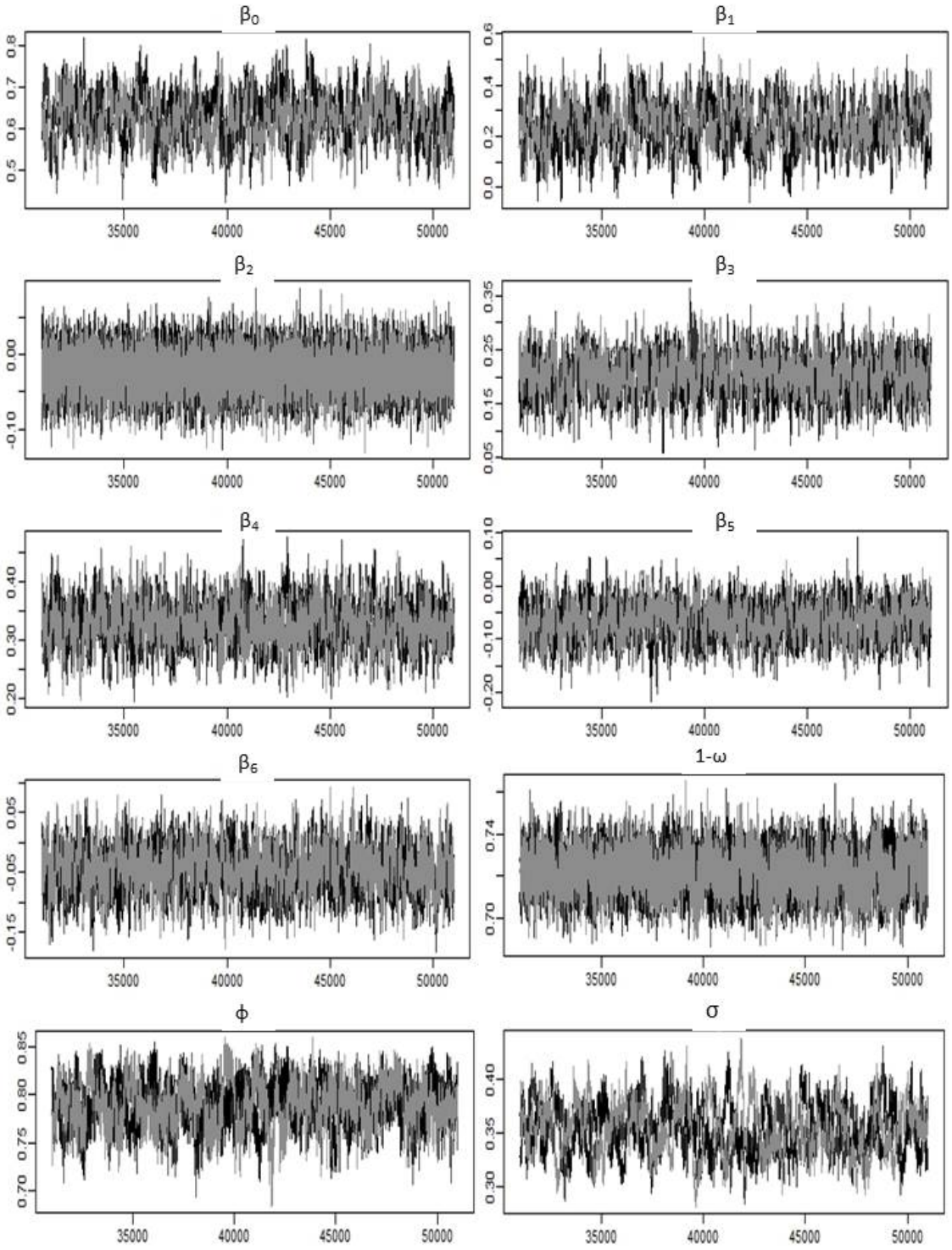


Figure 2.7: The trace plots of ZIP model parameters.



2.5.3 Asthma type J4591 dataset

We used a dataset of daily counts of emergency department visits of asthma type J4591 cases in Ontario, Canada during the period January 1st, 2012 till December 31st, 2016 (sample size = 1827). The dataset was obtained from the Canadian Institute for Health Information (CIHI). Daily counts that are below four were suppressed here, therefore, we translated the counts down by 4, as was done in the previous section. Figure 2.8 and 2.9 present, respectively, time series and histogram display of the data set.

Figure 2.8: Asthma presentation for asthma type J4591.

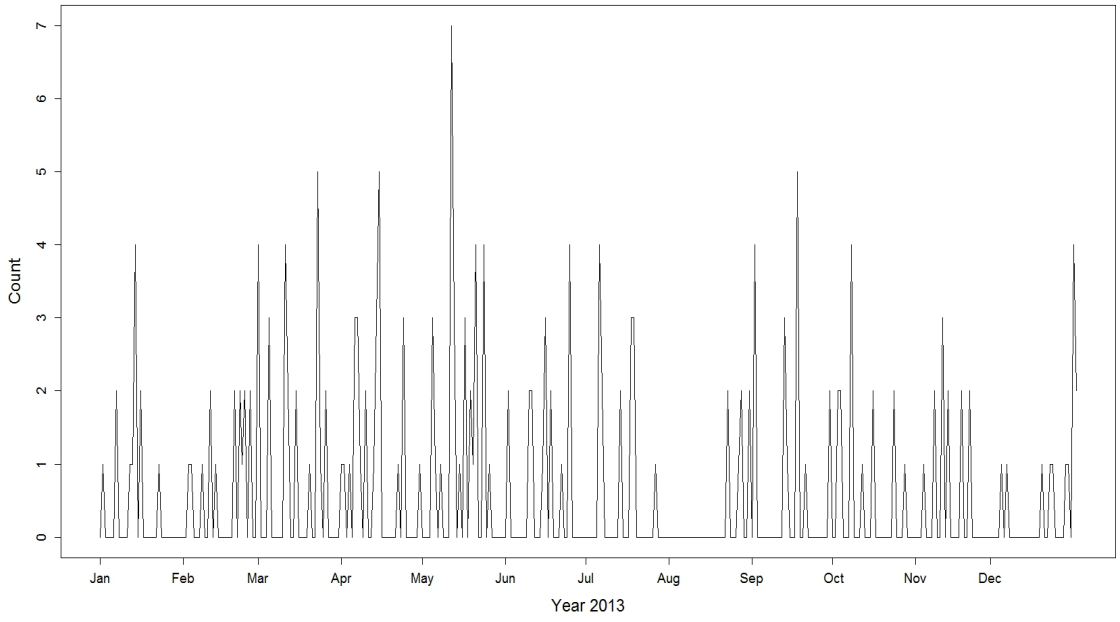
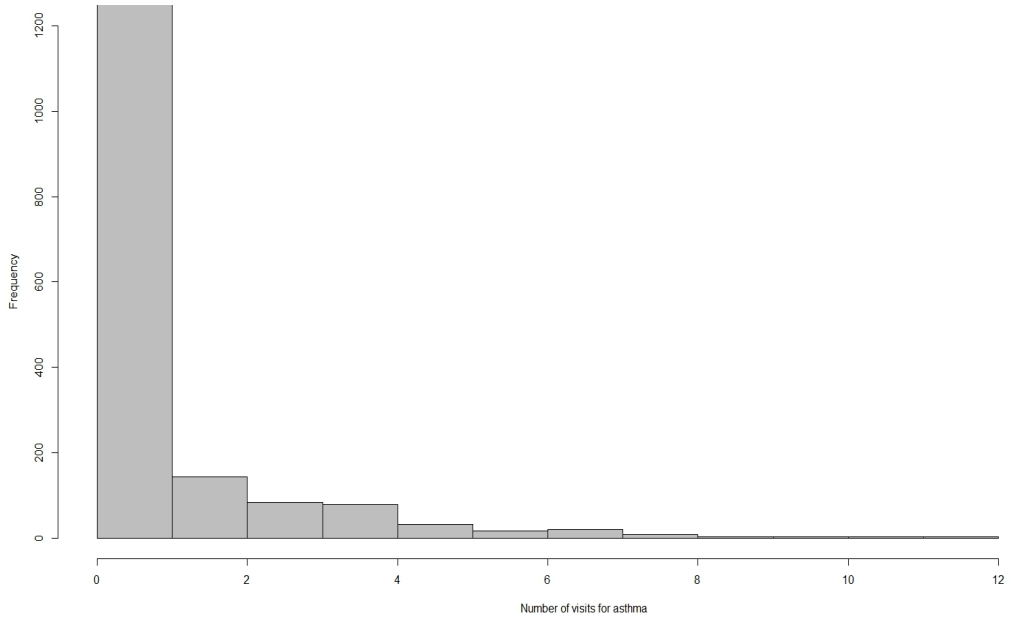


Figure 2.9: Histogram of daily visits of asthma type J4591 cases in Ontario. (January 2012 - December 2016)



The data were fitted using the following set of explanatory variables: $x_{1t} = \frac{t}{1827}$, to include time trend; x_{2t} to include weekend effect; $x_{3t} = cos\frac{2\pi t}{365}$; $x_{4t} = sin\frac{2\pi t}{365}$; $x_{5t} = cos\frac{4\pi t}{365}$ and $x_{6t} = sin\frac{4\pi t}{365}$ to include annual and semi-annual pattern. For weekend variable we coded it 1 for Saturday and Sunday and zero elsewhere. The dependent variable is emergency department daily asthma visits by people with asthma type J4591.

We fitted ZIP parameter driven model to asthma data assuming that there is AR(1) latent process, $\alpha_t = \phi_1\alpha_{t-1} + \epsilon_t$, where $\epsilon_t \sim Normal(0, \sigma^2)$, and presented the results in Table 2.14.

Table 2.14: Estimates and their standard errors from analysis of asthma data by ZIP parameter-driven model

Parameter	MLE	DC SE	z-value
β_0	0.2284	0.1260	1.81
β_1	0.4481	0.1596	2.81*
β_2	0.0385	0.0914	0.42
β_3	0.1008	0.0708	1.42
β_4	-0.0001	0.0678	0.00
β_5	-0.1099	0.0666	1.65
β_6	-0.0395	0.0743	-0.53
ω	0.5723	0.0207	27.65**
ϕ_1	0.4681	0.1599	2.93*
σ	0.6072	0.0715	8.49**

* indicates significant at 0.01 level ** indicates significant at 0.001 level

The models were fitted in *jags* and called into R using package *rjags*. We ran the sampler for 40000 iterations, discarding the first 20000 as burn-in and using three MCMC chains. Furthermore, the values of the maximum eigenvalue of the posterior variance-covariance matrix, mean square error, correlation-like fit statistic r^2 and Brooks-Gelman statistic, $\widehat{R}$, with K=6 are 0.0062, 0.0093, 0.0002 and 1.0291, respectively. These values reflect the degenerateness of the posterior distribution, the adequacy of the normal approximation and the convergence of MCMC chains. Figure 2.10 shows the posterior densities of ZIP parameter driven model which look appropriately normal, while Figure 2.11 presents trace plots of

ZIP parameter driven model, which suggested that samples of variance component show less quality of mixing than the samples of the other parameters.

From Table 2.14 we see that weekend, annual and semi-annual seasonal effects are not significant, while the trend is significant at 0.01 level. The variance component is statistically significant at 0.001 level, which means the number of asthma visitors to the emergency department of hospitals express significant heterogeneity. Also, significant correlation is detected at 0.001 level.

Figure 2.10: The posterior densities of ZIP model parameters.

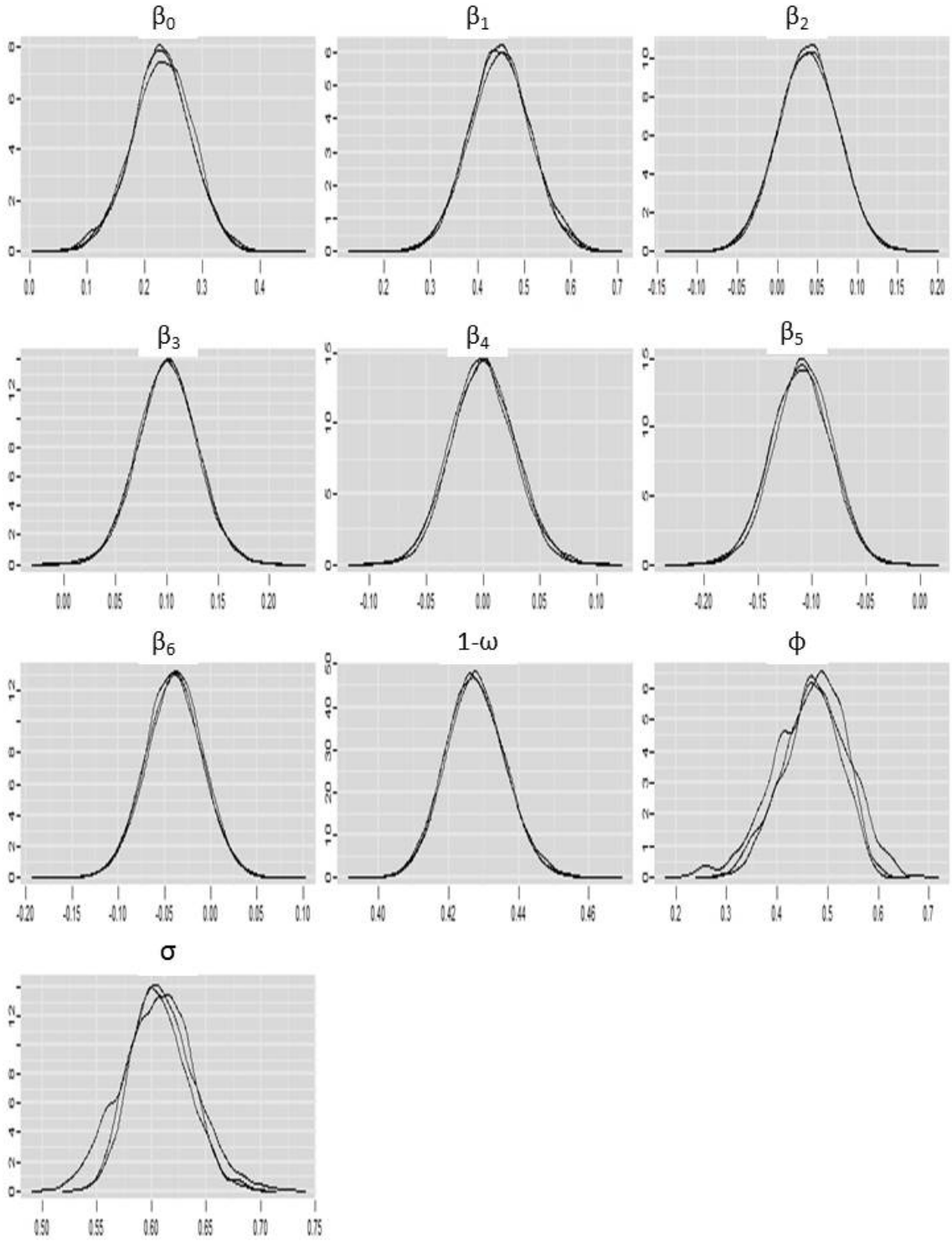
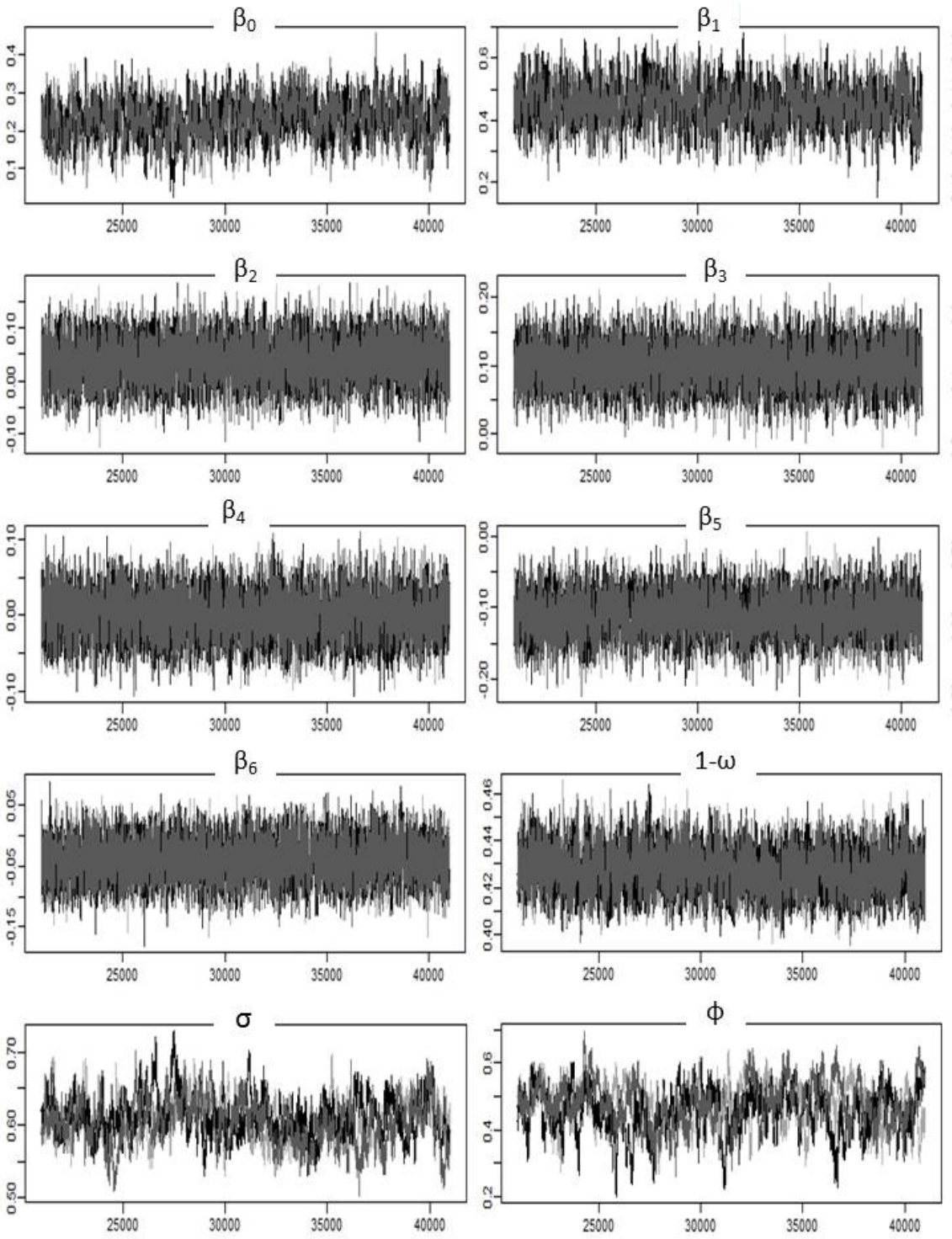


Figure 2.11: The trace plots of ZIP model parameters.



Chapter 3

Zero-inflated negative binomial parameter-driven model

This chapter proceeds as follows: in Section 1, we present a negative binomial parameter-driven model and its moment properties. To accommodate autocorrelations and excess zeros, we propose zero-inflated negative binomial (ZINB) parameter-driven model. In Section 2, MLEs for the model parameters are obtained via data cloning, and a simulation study is conducted in Section 3. Finally, real data applications are provided in Section 4.

3.1 Negative binomial regression models

Let $\{Y_t : t = 1, 2, \cdots, n\}$ be a time series of observed counts, $x_t^T = (x_{t_1}, \cdots, x_{t_k})$ is the tth row of covariate matrix X and $\beta = (\beta_1, \cdots, \beta_k)^T$ is the vector of regression coefficients of primary interest. Then a negative binomial regression model is

given by

$$f(y_t|x_t) = \begin{pmatrix} y_t + r - 1 \\ \\ y_t \end{pmatrix} p_t^r (1-p_t)^{y_t}, \qquad y_t = 0, 1, 2, \cdots$$

where r is a positive number and p_t satisfies the logit model

$$log\left\{\frac{r(1-p_t)}{p_t}\right\} = x_t^T\beta = \sum_{j=1}^{k} \beta_j x_{tj}.$$

Notice that under this representation, the mean and the variance are

$$EY_t = \frac{r(1-p_t)}{p_t} \quad \text{and} \quad Var(Y_t) = \frac{r(1-p_t)}{p_t^2}.$$

The zero-inflated negative binomial (ZINB) regression model has the distribution

$$f(y_t|x_t, z_t) = \begin{cases} \omega_t + (1-\omega_t)p_t^r, & \text{if } y_t = 0 \\ \\ (1-\omega_t) \begin{pmatrix} y_t + r - 1 \\ \\ y_t \end{pmatrix} p_t^r (1-p_t)^{y_t}, & \text{if } y_t > 0 \end{cases} \tag{3.1.1}$$

where

$$log\left\{\frac{r(1-p_t)}{p_t}\right\} = x_t^T\beta = \sum_{j=1}^{k} \beta_j x_{tj} \quad \text{and} \quad log\left(\frac{\omega_t}{1-\omega_t}\right) = z_t^T\gamma = \sum_{j=1}^{m} \gamma_j z_{tj},$$

$x_t^T = (x_{t_1}, \cdots, x_{t_k})$ is the tth row of covariate matrix X, $\beta = (\beta_1, \cdots, \beta_k)^T$ are unknown $k-$dimensional column vector of parameters, $z_t^T = (z_{t_1}, \cdots, z_{t_m})$ is the tth row of covariate matrix Z, $\gamma = (\gamma_1, \cdots, \gamma_m)^T$ are unknown $m-$dimensional column vector of parameters.

The mean and the variance of the ZIBN model [3.1.1] are

$$EY_t = r(1-\omega_t)\frac{(1-p_t)}{p_t} = (1-\omega_t)e^{x_t^T\beta},$$

and

$$\begin{aligned} Var(Y_t) &= \frac{r(1-\omega_t)(1-p_t)}{p_t^2}[1+r\omega_t(1-p_t)] \\ &= (1-\omega_t)e^{x_t^T\beta}[1+\omega_t e^{x_t^T\beta}+\frac{1}{r}e^{x_t^T\beta}]. \end{aligned}$$

3.1.1 Negative binomial parameter driven model

Consider a stationary autoregressive process of order p, $(AR(p))$, such that

$$\alpha_t = \phi_1\alpha_{t-1} + \phi_2\alpha_{t-2} + \cdots + \phi_p\alpha_{t-p} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a normal random process with mean zero and variance σ^2. Conditioning on α_t, assume that the random variables $Y_1, \cdots, Y_n$ are independent with negative binomial distribution. Specifically, consider the distribution

$$f(y_t|\alpha_t, x_t) = \begin{pmatrix} y_t+r-1 \\ y_t \end{pmatrix} p_t^r(1-p_t)^{y_t}, \qquad y_t = 0,1,2,\cdots \tag{3.1.2}$$

where r is a positive number and p_t satisfies

$$log\left\{\frac{r(1-p_t)}{p_t}\right\} = x_t^T\beta + \alpha_t,$$

where x_t and β are defined as mentioned before in negative binomial model. Also, assume that

$$f(y_t|\alpha_t) = f(y_t|\alpha_t, \boldsymbol{\alpha}^{(t-1)}) = f(y_t|\alpha_t, \boldsymbol{\alpha}^{(t-1)}, \mathbf{y}^{(t-1)}), \quad t = 1,2,\cdots \tag{3.1.3}$$

where $\mathbf{y}^{(t)} = (y_t, y_{t-1}, \cdots, y_1)$ and $\boldsymbol{\alpha}^{(t)} = (\alpha_t, \alpha_{t-1}, \cdots, \alpha_0, \alpha_{-1}, \cdots, \alpha_{1-p})$.

The first two moments of the observed process $\{Y_t\}$ can be evaluated as follows:

$$EY_t = E(E(Y_t|\alpha_t)) = E\left(\frac{r(1-p_t)}{p_t}\right) = E(e^{x_t^T\beta+\alpha_t}) = e^{x_t^T\beta}Ee^{\alpha_t} = e^{x_t^T\beta}\mu_\alpha,$$

where $\mu_\alpha = Ee^{\alpha_t}$.

$$\begin{aligned}
\sigma_{y_t}^2 = Var(Y_t) &= E(Var(Y_t|\alpha_t)) + Var(E(Y_t|\alpha_t)) \\
&= E\left[\frac{r(1-p_t)}{p_t^2}\right] + Var\left[\frac{r(1-p_t)}{p_t}\right] \\
&= E\left(e^{x_t^T\beta+\alpha_t}\left[1+\frac{1}{r}e^{x_t^T\beta+\alpha_t}\right]\right) + Var(e^{x_t^T\beta+\alpha_t}) \\
&= e^{x_t^T\beta}Ee^{\alpha_t} + \frac{1}{r}e^{2x_t^T\beta}E(e^{\alpha_t})^2 + e^{2x_t^T\beta}Var(e^{\alpha_t}) \\
&= e^{x_t^T\beta}\mu_\alpha + \frac{1}{r}\, e^{2x_t^T\beta}\mu_\alpha^2 + (1+\frac{1}{r})e^{2x_t^T\beta}\sigma_\alpha^2,
\end{aligned}$$

where $\sigma_\alpha^2 = Var(e^{\alpha_t})$.

The autocovariance function of Y_t is, for $h \neq 0$,

$$\begin{aligned}
\gamma_{y_t}(h) =& E(Cov(Y_t|\boldsymbol{\alpha}^{(t+h)}, Y_{t+h}|\boldsymbol{\alpha}^{(t+h)})) + Cov(E(Y_t|\boldsymbol{\alpha}^{(t+h)}), E(Y_{t+h}|\boldsymbol{\alpha}^{(t+h)})) \\
=& E(Cov(Y_t|\alpha_t, Y_{t+h}|\alpha_{t+h})) + Cov(E(Y_t|\alpha_t), E(Y_{t+h}|\alpha_{t+h})) \\
=& Cov(e^{x_t^T\beta+\alpha_t}, e^{x_{t+h}^T\beta+\alpha_{t+h}}) \\
=& e^{(x_t+x_{t+h})^T\beta}Cov(e^{\alpha_t}, e^{\alpha_{t+h}}) \\
=& e^{(x_t+x_{t+h})^T\beta}\gamma_\alpha(h),
\end{aligned}$$

where $\gamma_\alpha(h)$ is the autocovariance function of the latent process $\{e^{\alpha_t}\}$. Notice that the above derivation follows from proposition B.1. in Appendix B. Of course $\{Y_t\}$ is not a stationary time series since all its moments depend on t.

Suppose that $\{\alpha_t\}$ is an $AR(1)$ process, such that $\alpha_t = \phi\alpha_{t-1} + \epsilon_t$. Then the marginal moments of Y_t are obtained as follows:

$$EY_t = e^{\frac{\sigma^2}{2(1-\phi^2)}+x_t^T\beta},$$

$$\begin{aligned} Var(Y_t) =& e^{x_t^T\beta}e^{\frac{\sigma^2}{2(1-\phi^2)}} + \frac{1}{r}\ e^{2x_t^T\beta}e^{\frac{\sigma^2}{(1-\phi^2)}} + (1+\frac{1}{r})e^{2x_t^T\beta}e^{\frac{\sigma^2}{(1-\phi^2)}}\left[e^{\frac{\sigma^2}{(1-\phi^2)}}-1\right] \\ =& e^{\frac{\sigma^2}{2(1-\phi^2)}+x_t^T\beta}\left\{1+\frac{1}{r}\ e^{\frac{\sigma^2}{2(1-\phi^2)}+x_t^T\beta} + \left(\frac{1+r}{r}\right)e^{\frac{\sigma^2}{2(1-\phi^2)}+x_t^T\beta}\left[e^{\frac{\sigma^2}{(1-\phi^2)}}-1\right]\right\}, \end{aligned}$$

and for $h \neq 0$,

$$\gamma_{y_t}(h) = Cov(Y_t, Y_{t+h}) = e^{(x_t+x_{t+h})^T\beta+\frac{\sigma^2}{(1-\phi^2)}}\left[e^{\frac{\sigma^2\phi^h}{(1-\phi^2)}}-1\right].$$

Clearly, if $\phi > 0$, then the correlation between Y_t and Y_{t+h} is always positive.

3.1.2 ZINB parameter driven model

To accommodate correlation between successive observations in ZINB regression model, a latent process $\{\alpha_t\}$, is introduced in the link function. Specifically, let $\{\alpha_t\}$ be a stationary autoregressive process of order p, such that

$$\alpha_t = \phi_1\alpha_{t-1} + \phi_2\alpha_{t-2} + \cdots + \phi_p\alpha_{t-p} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a normal random process with mean zero and variance σ^2, conditional on the covariates and the latent process $\{\alpha_t\}$, $Y_1, \cdots, Y_n$ are independent and

modelled by a ZINB distribution, namely

$$f(y_t|\alpha_t, x_t) = \begin{cases} \omega + (1-\omega)p_t^r, & \text{if } y_t = 0 \\ (1-\omega)\begin{pmatrix} y_t + r - 1 \\ \\ y_t \end{pmatrix} p_t^r(1-p_t)^{y_t}, & \text{if } y_t > 0 \end{cases} \tag{3.1.4}$$

where

$$log\left\{\frac{r(1-p_t)}{p_t}\right\} = x_t^T\beta + \alpha_t,$$

$0 < \omega < 1, x_t$ and β are defined as mentioned before in ZINB model. Also, assume the validity of assumption 3.1.3. The marginal moments of the observed process $\{Y_t\}$ are given as follows:

$$EY_t = E(E(Y_t|\alpha_t)) = E\left\{(1-\omega)e^{x_t^T\beta+\alpha_t}\right\} = (1-\omega)e^{x_t^T\beta}Ee^{\alpha_t} = \mu_t\mu_\alpha,$$

where $\mu_t = (1-\omega)e^{x_t^T\beta}$ and $\mu_\alpha = Ee^{\alpha_t}$.

$$\begin{aligned} Var(Y_t) =& E(Var(Y_t|\alpha_t)) + Var(E(Y_t|\alpha_t)) \\ =& E\left\{(1-\omega)e^{x_t^T\beta+\alpha_t}(1+\omega e^{x_t^T\beta+\alpha_t} + \frac{1}{r}\ e^{x_t^T\beta+\alpha_t})\right\} + Var((1-\omega)e^{x_t^T\beta+\alpha_t}) \\ =& (1-\omega)e^{x_t^T\beta}\left\{Ee^{\alpha_t} + \omega e^{x_t^T\beta}E(e^{\alpha_t})^2 + \frac{1}{r}\ e^{x_t^T\beta}E(e^{\alpha_t})^2\right\} + (1-\omega)^2e^{2x_t^T\beta}Var(e^{\alpha_t}) \\ =& \mu_t\left\{\mu_\alpha + \frac{\omega}{1-\omega}\mu_t(\sigma_\alpha^2+\mu_\alpha^2) + \frac{1}{r(1-\omega)}\mu_t(\sigma_\alpha^2+\mu_\alpha^2)\right\} + \mu_t^2\sigma_\alpha^2 \\ =& \mu_t\mu_\alpha + \mu_t^2\mu_\alpha^2 + \frac{\omega}{1-\omega}\mu_t^2(\sigma_\alpha^2+\mu_\alpha^2) + \frac{1}{r(1-\omega)}\mu_t^2(\sigma_\alpha^2+\mu_\alpha^2), \end{aligned}$$

where $\sigma_\alpha^2 = Var(e^{\alpha_t})$.

$$\gamma_{y_t}(h) = E(Cov(Y_t|\boldsymbol{\alpha}^{(t+h)}, Y_{t+h}|\boldsymbol{\alpha}^{(t+h)})) + Cov(E(Y_t|\boldsymbol{\alpha}^{(t+h)}), E(Y_{t+h}|\boldsymbol{\alpha}^{(t+h)}))$$

$$
\begin{aligned}
&= E(Cov(Y_t|\alpha_t, Y_{t+h}|\alpha_{t+h})) + Cov(E(Y_t|\alpha_t), E(Y_{t+h}|\alpha_{t+h})) \\
&= Cov((1-\omega)e^{x_t^T\beta+\alpha_t}, (1-\omega)e^{x_{t+h}^T\beta+\alpha_{t+h}}) \\
&= (1-\omega)^2 e^{x_t^T\beta} e^{x_{t+h}^T\beta} Cov(e^{\alpha_t}, e^{\alpha_{t+h}}) \\
&= \mu_t \mu_{t+h} \gamma_\alpha(h),
\end{aligned}
$$

where $\gamma_\alpha(h)$ is the autocovariance function of the latent process $\{e^{\alpha_t}\}$.

3.2 Estimation

3.2.1 Parameter estimation of negative binomial model

To estimate the parameters of the negative binomial (NB) model [3.1.2], we need to write it as a hierarchical model. To proceed, consider the following model

$$Y_t|x_t, \alpha_t \backsim NB(r, p_t), \text{ with } \tfrac{r(1-p_t)}{p_t} = \exp(x_t^T\beta + \alpha_t),$$

$$\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p} \backsim \text{Normal}(\phi_1\alpha_{t-1} + \cdots + \phi_p\alpha_{t-p}, \sigma^2).$$

The likelihood function of this model is obtained by

$$L(\theta, y) = \int \prod_{t=1}^{n} f(y_t|\alpha_t) g(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p}) g_0(\alpha_0, \cdots, \alpha_{1-p})\, d\alpha,$$

for details see Appendix A.1. The posterior distribution of $\theta = (\beta, \phi_1, \cdots, \phi_p, \sigma)$ conditional on the data $y^{(K)} = (y, \cdots, y)$ is given by

$$\pi_K(\theta|y) = \frac{[L(\theta, y)]^K \pi(\theta)}{C(K, y)},$$

where $C(K, y) = \int [L(\theta, y)]^K \pi(\theta) d\theta$, is the normalizing constant.

3.2.2 Parameter estimation of ZINB model

In order to estimate the parameters of the ZINB model [3.1.4] using data cloning method, we need to write it as a hierarchical model. To this end, consider the following

$$Y_t|x_t, \alpha_t, u_t \backsim NB(r, p_t),$$

with

$$p_t = \frac{r}{r + (1 - u_t)\exp(x_t^T\beta + \alpha_t)} - 0.0000001u_t,$$

$$u_t \backsim \text{Bernoulli}(\omega),$$

$$\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p} \backsim \text{Normal}(\phi_1\alpha_{t-1} + \cdots, \phi_p\alpha_{t-p}, \sigma^2).$$

First, note that if $u_t = 1$, then $p_t \approx 1$,

$$P(Y_t = 0|\alpha_t, u_t = 1) = 1 \text{ and } P(Y_t = k|\alpha_t, u_t = 1) = 0, k > 0.$$

Second, if $u_t = 0$, then $p_t = \frac{r}{r+\exp(x_t^T\beta+\alpha_t)}$ and

$$P(Y_t = k|\alpha_t, u_t = 0) = \begin{pmatrix} k + r - 1 \\ k \end{pmatrix} p_t^r(1 - p_t)^k, \qquad k = 0, 1, 2, \cdots$$

Hence,

$$\begin{aligned} P(Y_t = 0|\alpha_t) &= P(u_t = 1)P(Y_t = 0|\alpha_t, u_t = 1) + P(u_t = 0)P(Y_t = 0|\alpha_t, u_t = 0) \\ &= \omega + (1 - \omega)p_t^r, \end{aligned}$$

and for $k > 0$,

$$P(Y_t = k|\alpha_t) = P(u_t = 1)P(Y_t = k|\alpha_t, u_t = 1) + P(u_t = 0)P(Y_t = k|\alpha_t, u_t = 0)$$

$$= (1-\omega)\begin{pmatrix} k+r-1 \\ k \end{pmatrix} p_t^r (1-p_t)^k,$$

which is the same distribution of Y_t in the model [3.1.4].

Now, the likelihood function of this model is given as follows:

$$L(\theta, y, \boldsymbol{\alpha}^{(n)}) = \prod_{t=1}^{n} f(y_t|\alpha_t) g(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p}) g_0(\alpha_0, \cdots, \alpha_{1-p}),$$

for details see Appendix A.1. Also, $f(y_t|\alpha_t) = \sum_{u_t=0}^{1} h(u_t) f(y_t|\alpha_t, u_t)$. Hence, the likelihood function of the model is

$$L(\theta, y) = \int \prod_{t=1}^{n} \left[\sum_{u_t=0}^{1} h(u_t) f(y_t|\alpha_t, u_t) \right] g(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p}) g_0(\alpha_0, \cdots, \alpha_{1-p}) \, d\alpha.$$

The posterior distribution of $\theta = (\omega, \beta, \phi_1, \cdots, \phi_p, \sigma)$ conditional on the data $y^{(K)} = (y, \cdots, y)$ is given by

$$\pi_K(\theta|y) = \frac{[L(\theta, y)]^K \, \pi(\theta)}{C(K, y)},$$

where $C(K, y) = \int [L(\theta, y)]^K \, \pi(\theta) d\theta$, is the normalizing constant.

3.3 Numerical studies

We considered two experiments: one with a negative binomial and the other with a ZINB density for the conditional distribution of observations given the latent process. For each case, we simulated 500 realizations and estimated the parameters of interest, reporting the empirical means, the empirical standard deviations and the mean square errors (MSE) of the estimates together with the

asymptotic standard deviation.

3.3.1 Experiment 1: Negative binomial.

Tables 3.1 and 3.3 show the real values of the parameters of model [3.1.2] from which the data has been simulated, the empirical means, the empirical standard deviations, MSE and DC standard errors with ($K = 3$). With the AR(1) latent process we used a covariate sequence defined by $x_t^T = (1, x_{t2})$, where x_{t2} is *uniform*(0,2) random variable, and with the AR(2) latent process we used a covariate sequence defined by $x_t^T = (1, sin\frac{\pi t}{6}, cos\frac{\pi t}{6})$, which includes two harmonic functions components. In both cases we set the sample size to be 500 and $r = 4$.

Tables 3.2 and 3.4 show the empirical means, the empirical standard deviations, MSE and Bayes MCMC standard errors with ($K = 1$).

The following priors were used in our simulations: normal distribution with mean zero and variance 10^4 for fixed effects parameters, $Gamma(1, 0.1)$ for the inverse of the variance component. For the parameter ϕ_1 of the latent process AR(1), we used $uniform(-0.99, 0.99)$, and for ϕ_1 and ϕ_2 in the latent process AR(2) we used normal prior distribution with mean zero and variance 10^4. Also, we set the normal prior distribution with mean zero and variance 1 for the initial condition parameters in both processes.

In each case we set the following: a burn-in period of 2000, two parallel MCMC chains and 5000 values to generate from the posterior distribution fron each chain.

Tables 3.1 and 3.2 show almost similar results of estimation for β_0 and β_1 and slightly better estimation for ϕ_1 and σ using DC method. In both tables the standard deviation estimates performed well. In Tables 3.3 and 3.4, we see that the empirical standard deviation is close to the estimated one and the estimates of the parameters are approximately unbiased with slightly better estimation for ϕ_2

using DC method. Table 3.5 reports the percentage of coverage of 95% confidence interval in both cases.

Table 3.1: Estimation of negative binomial model parameters with AR(1) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_0	-0.50	-0.5112	0.0174	0.1316	0.1299
β_1	0.90	0.9030	0.0108	0.1042	0.1034
σ	0.80	0.8048	0.0050	0.0703	0.0703
ϕ_1	-0.60	-0.5927	0.0045	0.0669	0.0647

Table 3.2: Estimation of negative binomial model parameters with AR(1) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_0	-0.50	-0.5161	0.0178	0.1326	0.1311
β_1	0.90	0.9046	0.0110	0.1050	0.1043
σ	0.80	0.8142	0.0051	0.0702	0.0707
ϕ_1	-0.60	-0.5852	0.0047	0.0672	0.0655

Table 3.3: Estimation of negative binomial model parameters with AR(2) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_0	0.50	0.4978	0.0050	0.0707	0.0740
β_1	-1.20	-1.2040	0.0110	0.1047	0.1028
β_2	2.00	2.0026	0.0125	0.1119	0.1105
σ	0.60	0.5974	0.0030	0.0547	0.0577
ϕ_1	1.00	0.9886	0.0034	0.0569	0.0551
ϕ_2	-0.75	-0.7434	0.0027	0.0512	0.0492

Table 3.4: Estimation of negative binomial model parameters with AR(2) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_0	0.50	0.4920	0.0051	0.0710	0.0749
β_1	-1.20	-1.2066	0.0111	0.1051	0.1048
β_2	2.00	2.0072	0.0127	0.1126	0.1118
σ	0.60	0.6133	0.0032	0.0548	0.0591
ϕ_1	1.00	0.9803	0.0037	0.0575	0.0568
ϕ_2	-0.75	-0.7323	0.0031	0.0528	0.0509

Table 3.5: Percentage of coverage of a 95% confidence interval in negative binomial model

Parameters	AR(1) latent process	AR(2) latent process
β_0	0.948	0.954
β_1	0.948	0.938
β_2	-	0.956
σ	0.948	0.952
ϕ_1	0.942	0.936
ϕ_2	-	0.936

3.3.2 Experiment 2: ZINB

In Tables 3.6 and 3.8 we show the real values of the parameters of model [3.1.4] from where the data has been simulated, the empirical means, the empirical standard deviations, MSE and DC standard errors with ($K = 3$). With both AR(1) and AR(2), we used a covariate sequence $x_t^T = (1, x_{t2})$, where x_{t2} is $uniform(0, 2)$ random variable and a sample size of 500.

Tables 3.7 and 3.9 show the empirical means, the empirical standard deviations, MSE and Bayes MCMC standard errors with ($K = 1$).

We set $Normal(0, 10^4)$ and $Gamma(1, 0.1)$ for β and the inverse of the variance component, respectively. Also, we used $uniform(-0.99, 0.99)$ For ϕ_1 in the latent process AR(1), and $Normal(0, 10^4)$ for ϕ_1 and ϕ_2 in the latent process AR(2). For the initial condition parameters in both processes we used $Normal(0, 1)$. The burn-in period was 5000, three parallel MCMC chains were generated with 10000

iterations in each one.

Tables 3.6 and 3.7 show almost similar results of estimation for the parameters β_0, β_1 and ω, and slightly better estimation for the latent process parameters ϕ_1 and σ using DC method, the standard deviation estimates performed well. Simulation results reported in Tables 3.8 and 3.9 show that the empirical standard deviation is close to the estimated one in both $K = 1$ and $K = 3$. In addition, we can see some bias in estimating ϕ_1 and ϕ_2 in both methods with slightly better estimation using DC method. Table 3.10 reports the percentage of coverage of a 95% confidence interval with both AR(1) and AR(2) processes.

Table 3.6: Estimation of ZINB model parameters with AR(1) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_0	0.50	0.4749	0.0205	0.1410	0.1396
β_1	1.50	1.5160	0.0118	0.1074	0.1080
ω	0.30	0.2854	0.0003	0.0116	0.0243
σ	0.70	0.7010	0.0048	0.0697	0.0700
ϕ_1	-0.70	-0.6907	0.0034	0.0580	0.0601

Table 3.7: Estimation of ZINB model parameters with AR(1) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_0	0.50	0.4692	0.0209	0.1415	0.1411
β_1	1.50	1.5184	0.0119	0.1075	0.1092
ω	0.30	0.2852	0.0004	0.0115	0.0243
σ	0.70	0.7138	0.0051	0.0701	0.0712
ϕ_1	-0.70	-0.6803	0.0040	0.0596	0.0619

Table 3.8: Estimation of ZINB model parameters with AR(2) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_0	-0.50	-0.5065	0.0159	0.0996	0.1348
β_1	2.00	2.0083	0.0089	0.0752	0.0972
ω	0.20	0.1990	0.0002	0.0105	0.0241
σ	0.60	0.5838	0.0046	0.0513	0.0654
ϕ_1	-0.40	-0.4540	0.0221	0.1088	0.1407
ϕ_2	0.45	0.3891	0.0224	0.1077	0.1400

Table 3.9: Estimation of ZINB model parameters with AR(2) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_0	-0.50	-0.5097	0.0160	0.0991	0.1359
β_1	2.00	2.0099	0.0089	0.0750	0.0978
ω	0.20	0.1996	0.0002	0.0104	0.0240
σ	0.60	0.5893	0.0042	0.0501	0.0649
ϕ_1	-0.40	-0.4609	0.0181	0.0969	0.1343
ϕ_2	0.45	0.3782	0.0193	0.0960	0.1334

Table 3.10: Percentage of coverage of a 95% confidence interval in ZINB model

Parameters	AR(1) latent process	AR(2) latent process
β_0	0.948	0.954
β_1	0.948	0.948
ω	1.00	1.00
σ	0.948	0.940
ϕ_1	0.960	0.948
ϕ_2	-	0.946

3.4 Application to real data

3.4.1 Polio dataset

We revisited polio data example again and used the same regression variables we used before in the Poisson model case, namely

$$x_t^T = (1, \tilde{t}/1000, cos(2\pi\tilde{t}/12), sin(2\pi\tilde{t}/12), cos(2\pi\tilde{t}/6), sin(2\pi\tilde{t}/6)),$$

where $\tilde{t} = t - 73$. We assumed that the latent process $\{\alpha_t\}$ is an AR(1) and the observations $y_t, (t = 1, \cdots, 168)$ conditional on the latent process are independent and following negative binomial distribution with mean

$$\frac{r(1-p_t)}{p_t} = exp(x_t^T\beta + \alpha_t).$$

For parameter estimation we adapted the following strategy. First, we obtained the MLEs for β, ϕ and σ with fixe values of $r, (r = 1, \cdots, 9)$, and number of colons=50. Secondly, we generated 500 random data samples of the latent process and hence calculated the AIC differences for all pairs of models with different values of r, then we chose the estimate $\hat{r} = 2$ that yielded positive AIC difference with respect to all the other values of r. Table 3.11 shows the results of the AIC differences.

Table 3.11: The AIC difference with $r = 2$ in model 2

r value in model 1	$AIC_1 - AIC_2$
1	50.70
3	0.32
4	9.78
6	31.56
7	40.66
9	58.14

Table 3.12 shows the estimates of β, ϕ and σ and their standard errors using 200 colons, three parallel MCMC chains for 30000 iterations each, following burn-in period of 5000 iterations. For comparison, we included the results from Poisson parameter driven model of Chapter 1. We used the AIC difference for model comparison, taking the first model to be Poisson and the second model to be negative binomial, the value of $AIC_1 - AIC_2 = 123.3$. Because the difference is positive and greater than 3 we conclude that negative binomial model provides better description of the data than does Poisson model. Both models suggest no evidence of a decreasing trend in the rate of US polio ifections over time.

To check the convergence of the DC approach with negative binomial model, in Table 3.13 we computed the largest eigenvalue of the posterior variance-covariance matrix (Lamda.max), mean squared error (ms.error)and correlation-like fit statistic (r^2) and all these statistics converge to zero. For further investigation of the behaviour of MCMC chains, we computed Brooks-Gelman statistic, $\widehat{R}$, and it converges to 1 as shown in Table 3.13. Figure 3.1 presents trace plots for the model

parameters and Figure 3.2 shows the posterior distribution for the parameters with ($K = 200$). Monitoring the trace plots suggested that the samples of the fixed effects and the samples of correlation parameter got mixed well, but the samples of the variance component shows less quality of mixing. The posterior densities of all the parameters look appropriately normal.

Table 3.12: Estimates and their standard errors from analysis of polio data by negative binomial and Poisson parameter driven models

Parameter	Poisson			Negative Binomial		
	MLE	DC SE	z-value	MLE	DC SE	z-value
β_0	-0.0308	0.1543	-0.20	0.1053	0.1854	0.57
β_1	-3.8600	2.8522	-1.35	-3.8976	3.7002	-1.05
β_2	-0.0974	0.1491	-0.65	-0.1117	0.1400	-0.80
β_3	-0.4954	0.1583	-3.13**	-0.5021	0.1499	-3.35**
β_4	0.1986	0.1249	1.59	0.1852	0.1378	1.34
β_5	-0.3627	0.1256	-2.89**	-0.3702	0.1356	-2.73**
ϕ_1	0.6561	0.1646	3.99**	0.8836	0.0884	10.00**
σ	0.5249	0.1235	4.25**	0.2290	0.0948	2.42*

* indicates significant at 0.05 level ** indicates significant at 0.01 level

Table 3.13: Estimability diagnostics

Number of clones	lambda.max	ms error	r^2	$\widehat{R}$
1	25.0997	8.4090	0.1273	1.0230
50	0.2711	0.0021	0.0001	1.0144
100	0.1314	0.0008	0.0000	1.0153
150	0.0995	0.0011	0.0000	1.0143
200	0.0685	0.0013	0.0001	1.0127

Figure 3.1: The trace plots of NB model parameters

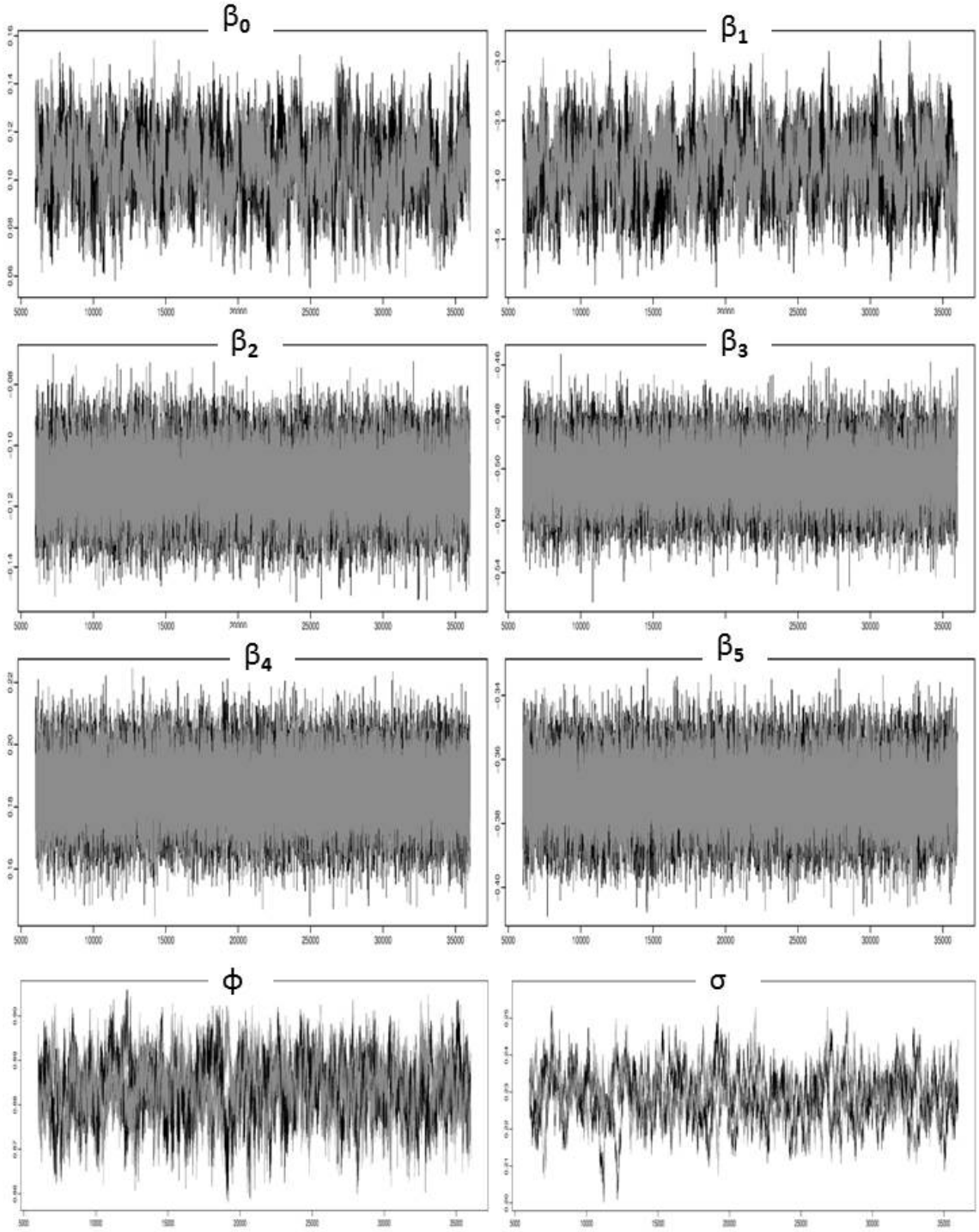
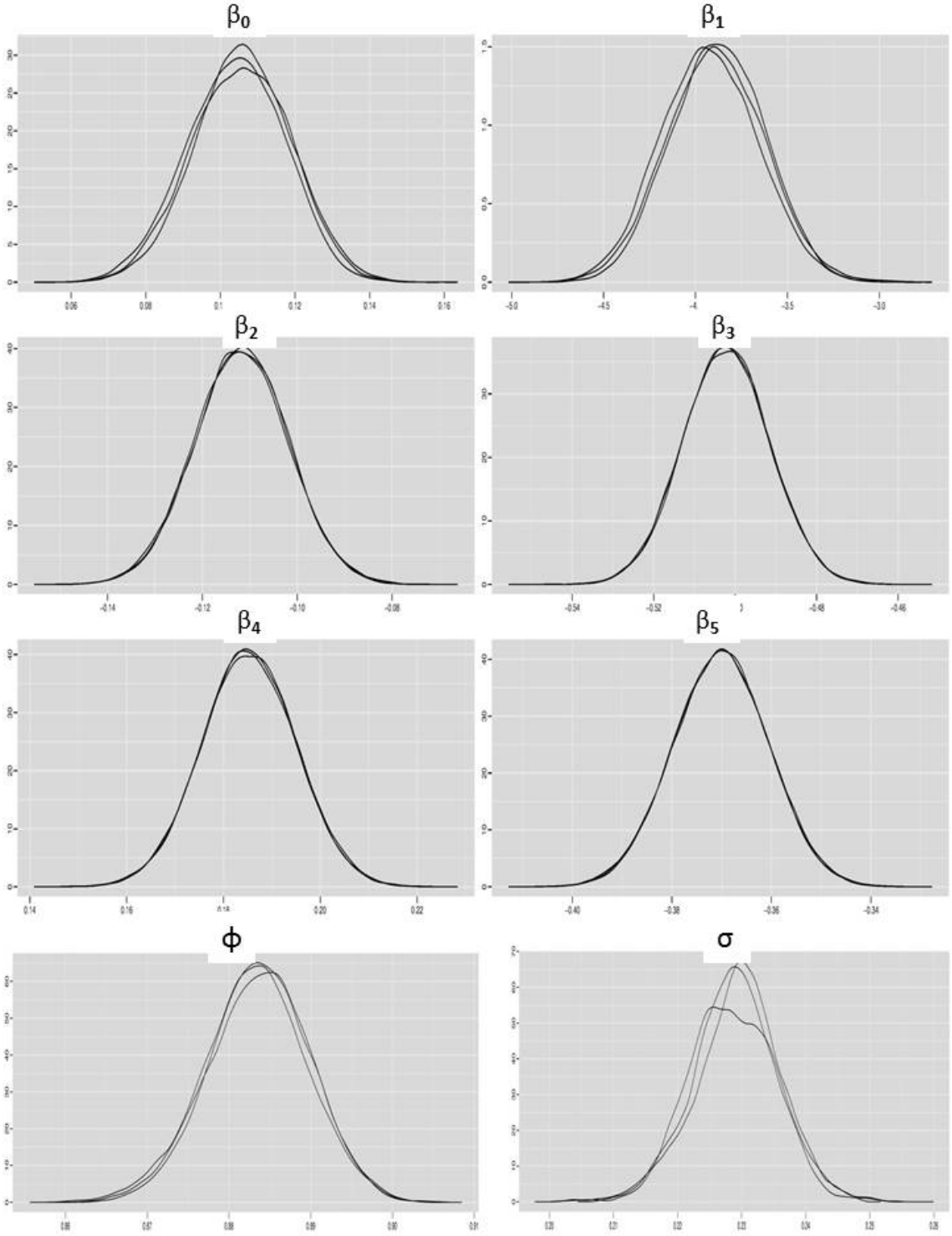


Figure 3.2: The posterior densities of NB model parameters



3.4.2 Asthma age group (70-79) years dataset

We revisited the dataset of daily counts of emergency department visits of asthma cases of people aged (70-79) years in Ontario, Canada again, and used the same regression variables we used before in ZIP model, namely: $x_{1t} = \frac{t}{2191}$; $x_{2t} =$ weekend; $x_{3t} = cos\frac{2\pi t}{365}$; $x_{4t} = sin\frac{2\pi t}{365}$; $x_{5t} = cos\frac{4\pi t}{365}$ and $x_{6t} = sin\frac{4\pi t}{365}$. We assumed that the latent process $\{\alpha_t\}$ is an AR(1) and the observations $y_t, (t = 1, \cdots, 2191)$, conditional on the latent process are independent and following ZINB distribution with mean $\frac{r(1-p_t)}{p_t} = exp(x_t^T\beta + \alpha_t)$.

For parameter estimation we adapted the following approach. First, find the MLE for β, ϕ and σ with different values of $r, (r = 1, 2, \cdots, 9)$, and number of clones =3. Second, generate 100 random data samples of the latent process. Third, calculate AIC difference for all pairs of models with different values of r. Finally, the estimate $\widehat{r} = 2$ was determined by the r value that yielded positive AIC difference with all the other models. Table 3.14 shows the results of the AIC difference.

Table 3.14: The AIC difference with $r = 2$ in model 2

r value in model 1	$AIC_1 - AIC_2$
1	52.52
3	8.86
4	312.39
5	440.98
6	736.83
9	1245.64

Table 3.15 presents MLE of ZINB parameter driven model using number of colons=5, three parallel MCMC chains with 30000 iterations each following burn-in period of 20000 iterations. For comparison, we included the results from ZIP parameter driven model of Chapter 1. The value of AIC difference was, $AIC_1 - AIC_2 = 3209.1$ assuming that ZIP is model 1 and ZINB is model 2, the value of AIC difference indicating that ZINB provides better description of the data than does ZIP model.

Table 3.15: Estimates and their standard errors from analysis of asthma data by ZIP and ZINB parameter driven models

Parameter	ZIP parameter driven model			ZINB parameter driven model		
	MLE	DC SE	z-value	MLE	DC SE	z-value
β_0	0.6238	0.0919	6.79**	0.5500	0.1234	4.46**
β_1	0.2445	0.1451	1.69	0.2630	0.2067	1.27
β_2	-0.0224	0.0446	-0.50	-0.0100	0.0568	-0.18
β_3	0.2015	0.0606	3.33**	0.2117	0.0793	2.67**
β_4	0.3278	0.0613	5.35**	0.3638	0.0797	4.56**
β_5	-0.0681	0.0595	-1.14	-0.0786	0.0721	1.09
β_6	-0.0477	0.0602	-0.79	-0.0510	0.0790	-0.65
ω	0.2784	0.0161	17.29**	0.1741	0.0182	9.57**
ϕ_1	0.7906	0.0364	21.72**	0.9275	0.0180	51.53**
σ	0.3520	0.0368	9.56**	0.1683	0.0254	6.63**

** indicates significant at 0.01 level

Table 3.15 shows that the estimates and standard errors in ZIP and ZINB

models are comparable and both suggest no evidence of increasing trend in the rate of asthma over time for the age group (70-79) years. In both models, annual seasonal effects and zero state probability are significant while semi-annual and weekend effects are not significant. The variance component in both models is statistically significant at 0.001 level, which means the number of asthma visitors to the emergency department of hospitals for the age group (70-79) years express significant heterogeneity. Also, significant correlation is detected at 0.001 level in both models with larger value in ZINB.

To check the convergence of the DC method with ZINB model, we calculated the largest eigenvalue of the posterior variance-covariance matrix, mean square error and correlation-like fit statistic and their values were 0.0110, 0.0126 and 0.0005, respectively, indicating the convergence of the approach. Also, we computed Brooks-Gelman statistic, $\widehat{R}$, and it was 1.0560 indicting the convergence of the chains. Furthermore, trace plots and posterior densities of the model parameters are shown in Figure 3.3 and 3.4, respectively. Monitoring the trace plots suggested that the samples of all parameters got mixed well except the samples of the variance component which shows less quality of mixing. The posterior densities of all the parameters look appropriately normal.

Figure 3.3: The trace plots of ZINB model parameters

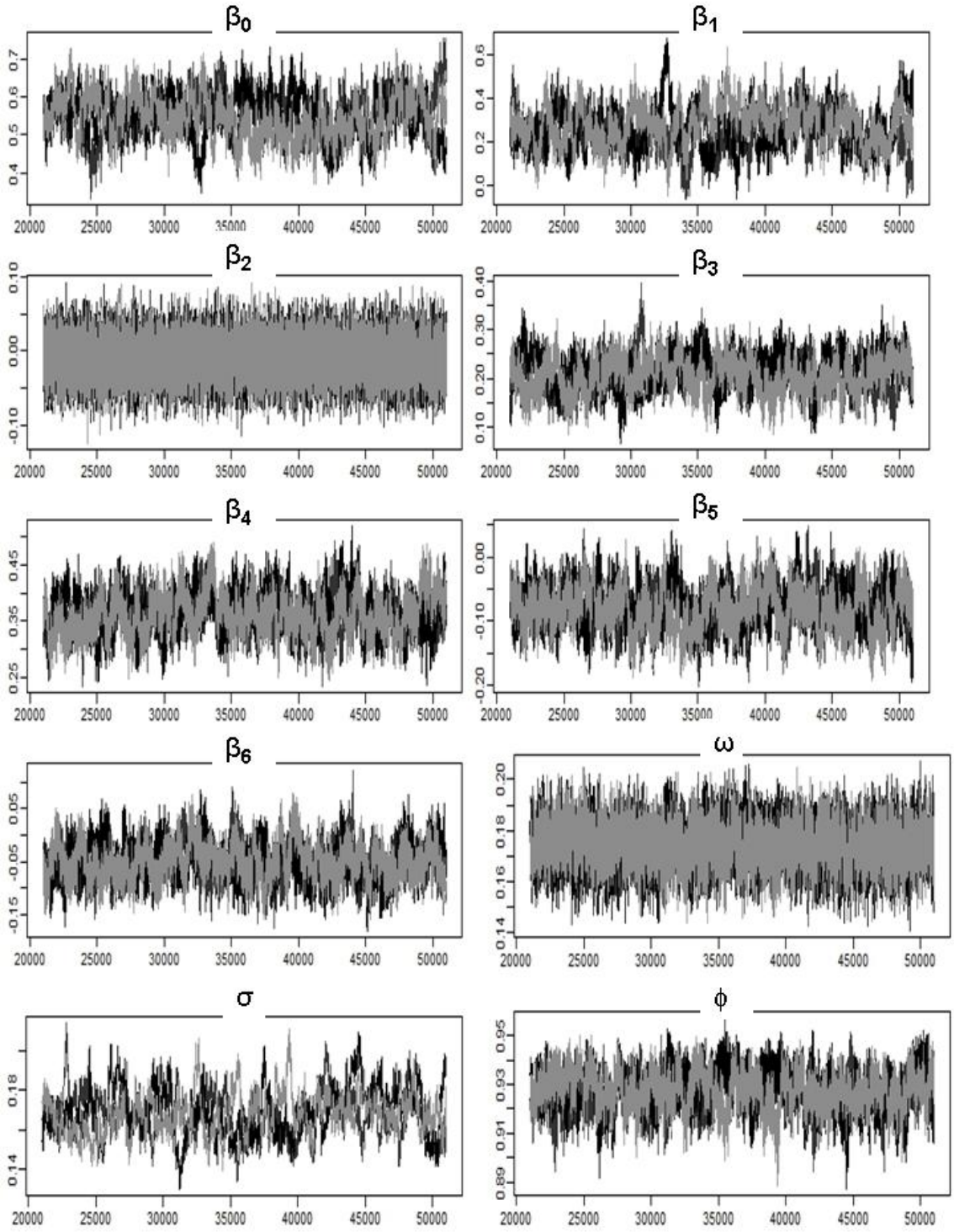
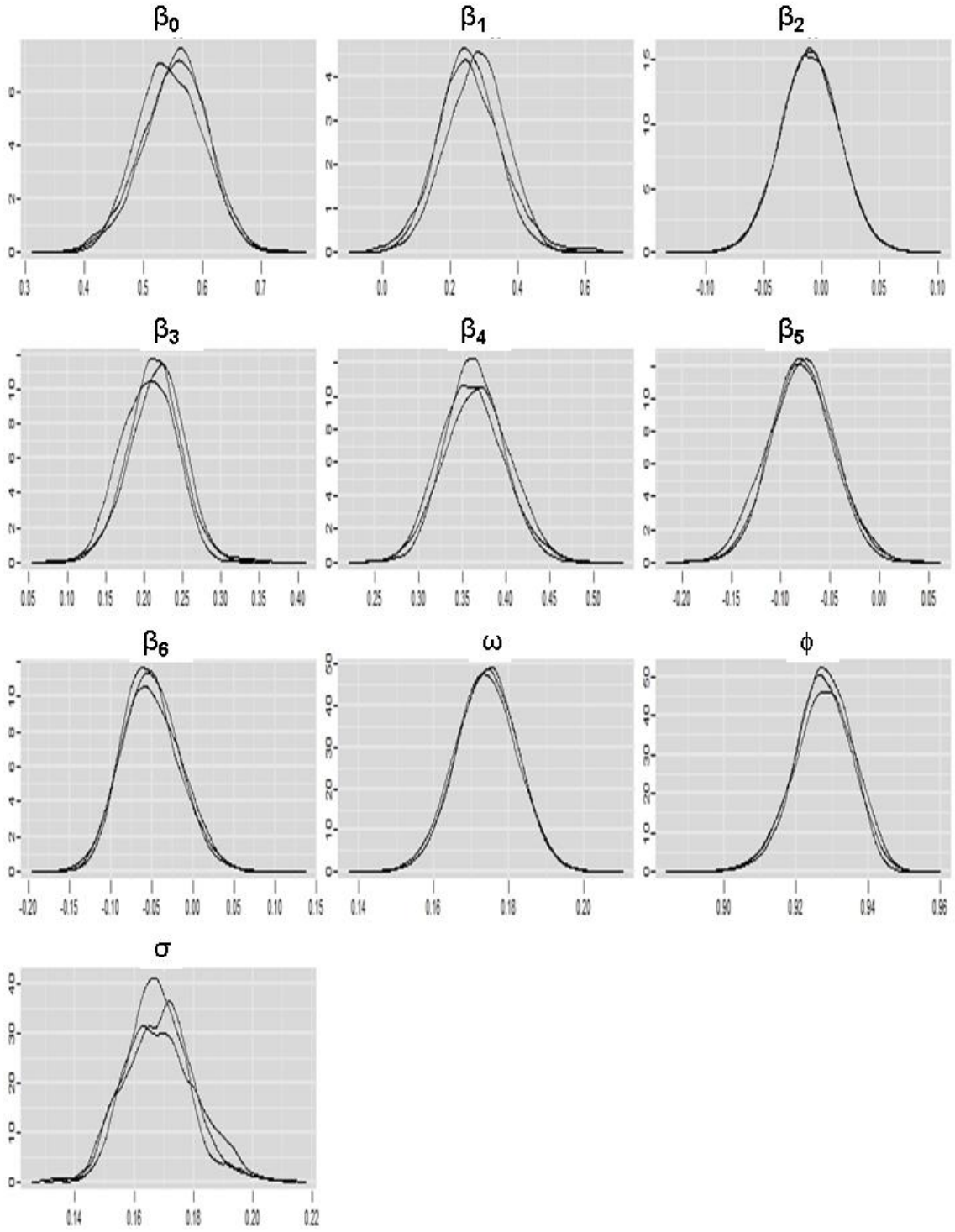


Figure 3.4: The posterior densities of ZINB model parameters



3.4.3 Asthma type J4591 dataset

We revisited the dataset of daily counts of emergency department visits of asthma type J4591 cases in Ontario, Canada again, and used the same regression variables we used before in ZIP model, namely: $x_{1t} = \frac{t}{1827}$; $x_{2t} =$ weekend; $x_{3t} = cos\frac{2\pi t}{365}$; $x_{4t} = sin\frac{2\pi t}{365}$; $x_{5t} = cos\frac{4\pi t}{365}$ and $x_{6t} = sin\frac{4\pi t}{365}$. We assumed that the latent process $\{\alpha_t\}$ is an AR(1) and the observations $y_t, (t = 1, \cdots, 1827)$, conditional on the latent process are independent and following ZINB distribution with mean

$$\frac{r(1-p_t)}{p_t} = exp(x_t^T\beta + \alpha_t).$$

For parameter estimation we adapted the following approach. First, find the MLE for β, ϕ and σ with different values of $r, (r = 1, 2, \cdots, 9)$, and number of clones =3. Second, generate 100 random data samples of the latent process. Third, calculate AIC difference for all pairs of models with different values of r. Finally, the estimate $\widehat{r} = 2$ was determined by the r value that yielded positive AIC difference with all the other models. Table 3.16 shows the results of the AIC difference.

Table 3.16: The AIC difference with $r = 2$ in model 2

r value in model 1	$AIC_1 - AIC_2$
1	74.40
3	677.92
4	566.48
5	1120.91
6	1456.12
7	2021.30

Table 3.17 presents MLE of ZINB parameter driven model using number of clones=5, three parallel MCMC chains with 20000 iterations each following burn-in period of 20000 iterations. For comparison, we included the results from ZIP parameter driven model of Chapter 1. The value of AIC difference was, $\text{AIC}_1 - \text{AIC}_2 = 6000.5$ assuming that ZIP is model 1 and ZINB is model 2, the value of AIC difference indicating that ZINB provides better description of the data than does ZIP model.

Table 3.17: Estimates and their standard errors from analysis of asthma data by ZIP and ZINB parameter driven models

Parameter	ZIP parameter driven model			ZINB parameter driven model		
	MLE	DC SE	z-value	MLE	DC SE	z-value
β_0	0.2284	0.1260	1.81	0.1980	0.1769	1.12
β_1	0.4481	0.1596	2.81*	0.5369	0.2913	1.84
β_2	0.0385	0.0914	0.42	0.0873	0.0917	0.95
β_3	0.1008	0.0708	1.42	0.0849	0.1092	0.78
β_4	-0.0001	0.0678	0.00	-0.0101	0.1166	-0.09
β_5	-0.1099	0.0666	1.65	-0.1313	0.0917	-1.43
β_6	-0.0395	0.0743	-0.53	-0.0584	0.1047	-0.56
ω	0.5723	0.0207	27.65**	0.5305	0.0194	27.35**
ϕ_1	0.4681	0.1599	2.93*	0.9571	0.0188	50.91**
σ	0.6072	0.0715	8.49**	0.1275	0.0328	3.89**

* indicates significant at 0.01 level ** indicates significant at 0.001 level

Table 3.17 shows that the estimates and standard errors in ZIP and ZINB models are comparable and both suggest no evidence of seasonal effect over time for type J4591. The variance component in both models is statistically significant at 0.001 level, which means the number of asthma visitors to the emergency department of hospitals of type J5491 express significant heterogeneity. Also, significant correlation is detected at 0.001 level in both models with larger value in ZINB. In ZIP model, there is significant increasing trend at level (0.01) of significance, while in ZINB model the trend in the rate of asthma over time was not significant.

To check the convergence of the DC method with ZINB model, we calculated the largest eigenvalue of the posterior variance-covariance matrix, mean square error and correlation-like fit statistic and their values were 0.0110, 0.0126 and 0.0005, respectively, indicating the convergence of the approach. Also, we computed Brooks-Gelman statistic, $\widehat{R}$, and it was 1.0560 indicting the convergence of the chains. Furthermore, trace plots and posterior densities of the model parameters are shown in Figure 3.5 and 3.6, respectively. Monitoring the trace plots suggested that the samples of all parameters got mixed well except the samples of the variance component which shows less quality of mixing. The posterior densities of all the parameters look appropriately normal.

Figure 3.5: The trace plots of ZINB model parameters

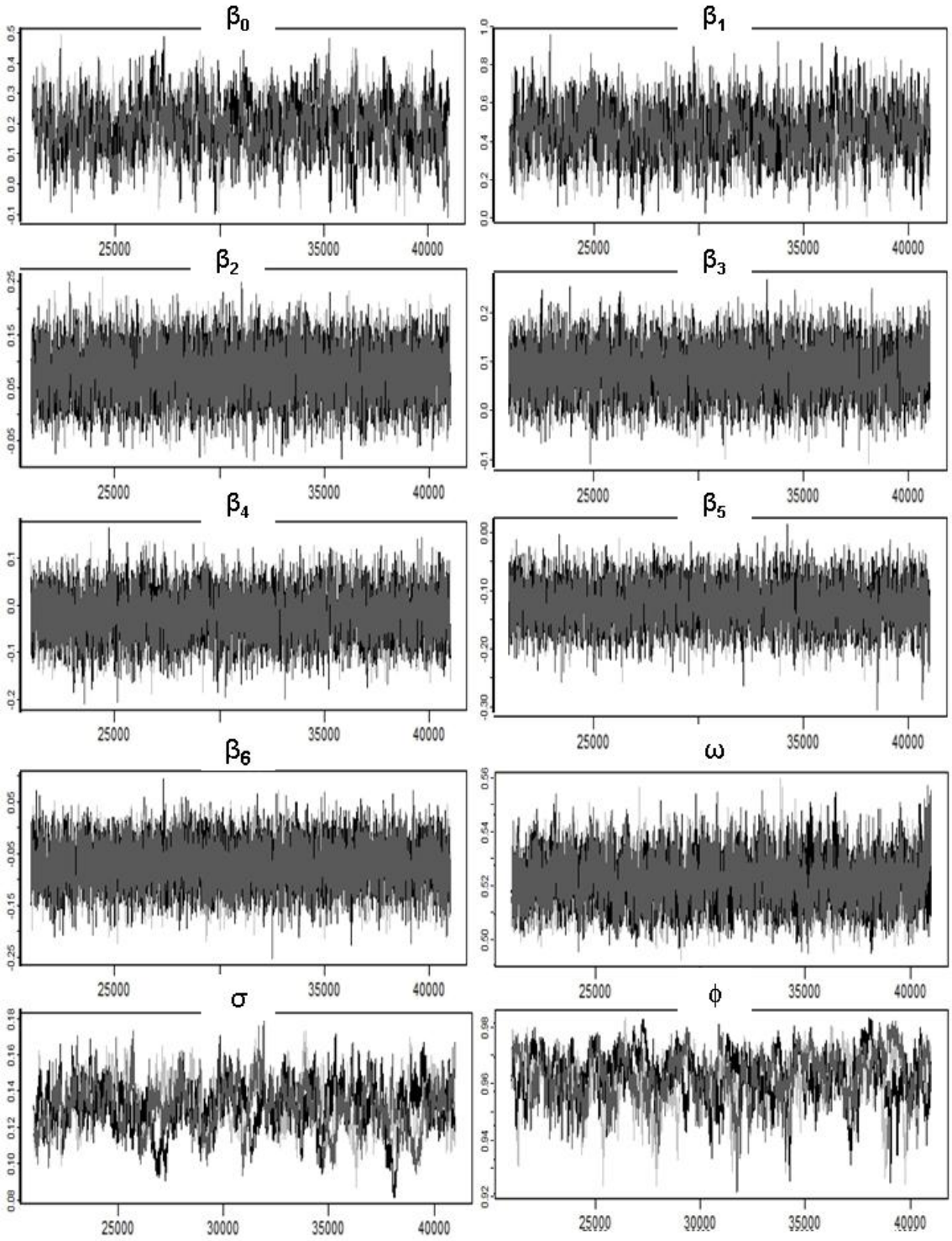
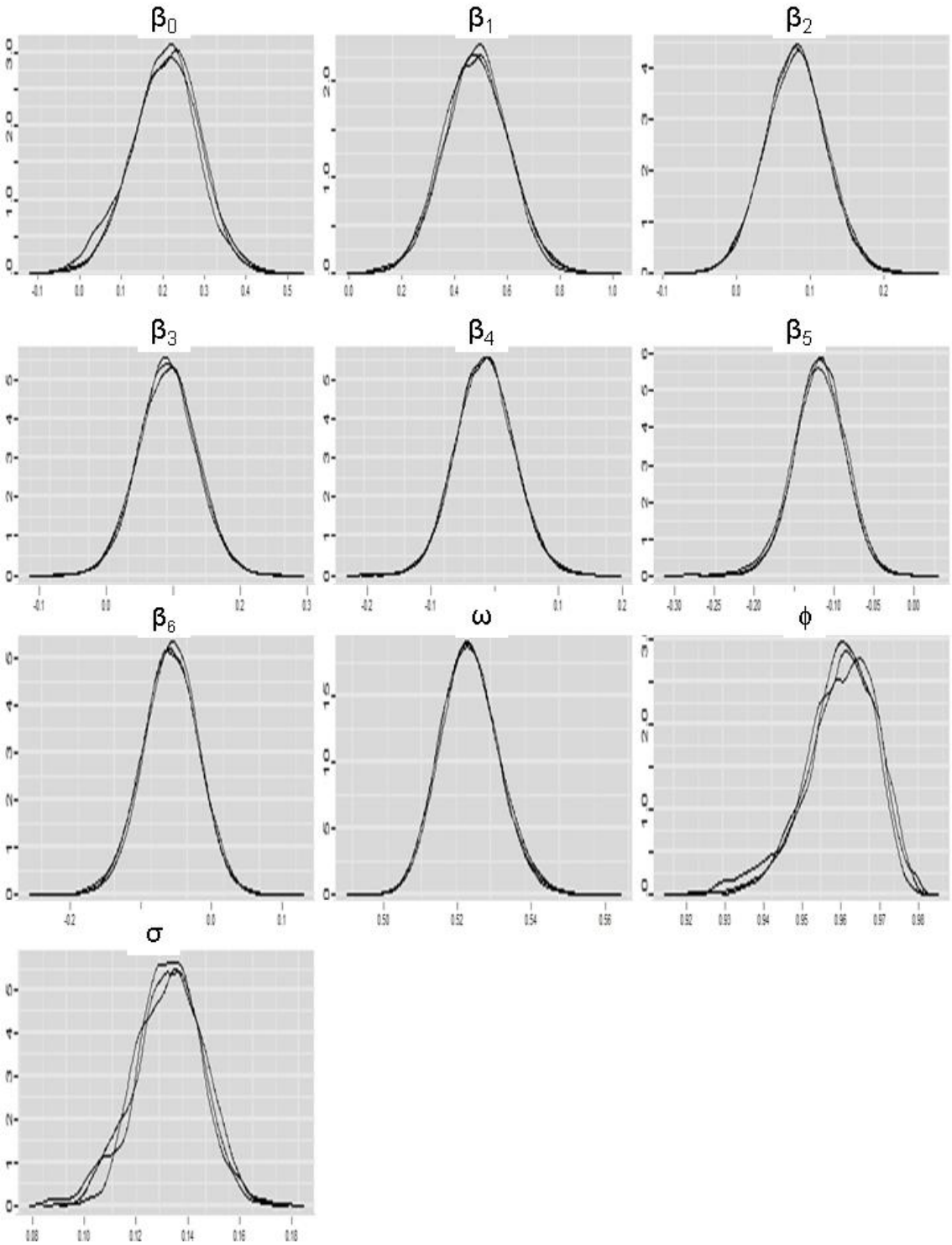


Figure 3.6: The posterior densities of ZINB model parameters



Chapter 4

Hurdle parameter-driven model

This chapter proceeds as follows: in Section 1 we present the hurdle Poisson and negative binomial parameter-driven models. Adaptation of the data cloning algorithm to estimate these parameters of these models is outlined in Section 2. A simulation study is conducted in Section 3, and finally, an application on a data set on emergency department visits due to asthma in the Canadian province of Ontario is discussed in Section 4.

4.1 Hurdle model

Assume that f_1 and f_2 are any probability density functions for non negative integers. Let $\{Y_t : t = 1, 2, \cdots, n\}$ be a time series of observed counts. Then a hurdle model can be expressed as:

$$f(Y_t = y_t) = \begin{cases} f_1(0), & \text{if } y_t = 0 \\ (1 - f_1(0))\frac{f_2(y_t)}{1-f_2(0)}, & \text{if } y_t > 0 \end{cases} \tag{4.1.1}$$

The model collapses to the standard model if $f_1(0) = f_2(0)$, allows for excess zeros if $f_1(0) > f_2(0)$ and too few zeros if $f_1(0) < f_2(0)$.

The moments of the model are determined by the probability of crossing the threshold and by the moments of the zero-truncated density. Namely, the mean is

$$EY_t = \frac{1 - f_1(0)}{1 - f_2(0)}\mu_2,$$

where μ_2 is the untruncated mean in density $f_2(y_t)$, and the variance is

$$var(Y_t) = \frac{1 - f_1(0)}{1 - f_2(0)}\sigma_2^2 + \frac{(1 - f_1(0))(f_1(0) - f_2(0))}{(1 - f_2(0))^2}\mu_2^2,$$

where σ_2^2 is the untruncated variance in density $f_2(y_t)$.

To extend the hurdle model to accommodate correlations in time series counts, we propose hurdle model with latent process $\{\alpha_t\}$ to handle such correlation in the following subsections.

4.1.1 Poisson autoregressive hurdle model

Let $\{Y_t : t = 1, 2, \cdots, n\}$ be a time series of observed counts, $x_t^T = (x_{t1}, \cdots, x_{tk})$ is the tth row of covariate matrix X, $\beta = (\beta_1, \cdots, \beta_k)^T$ are unknown k-dimensional column vector of parameters, $z_t^T = (z_{t1}, \cdots, z_{tm})$ is the tth row of the covariate matrix Z and $\gamma = (\gamma_1, \cdots, \gamma_m)^T$ are unknown m-dimensional column vector of parameters.

To accommodate correlation between successive observations of the time series when $f_2(\cdot)$ in Model [4.1.1] is specified as Poisson distribution, consider the following model: let $\{\alpha_t\}$ be a stationary autoregressive process of order p, AR(p),

such that

$$\alpha_t = \phi\alpha_{t-1} + \phi_2\alpha_{t-2} + \cdots + \phi_p\alpha_{t-p} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a normal random process with mean zero and variance σ^2. Conditioning on α_t, suppose Y_t is a sequence of independent counts with Poisson hurdle (PH) distribution defined as follows:

$$f(y_t|\alpha_t, x_t, z_t) = \begin{cases} \omega_t, & \text{if } y_t = 0 \\ (1-\omega_t)\frac{e^{-\lambda_t}\lambda_t^{y_t}}{(1-e^{-\lambda_t})y_t!}, & \text{if } y_t > 0 \end{cases} \tag{4.1.2}$$

where

$$log(\lambda_t) = x_t^T\beta + \alpha_t,$$

and

$$logit(\omega_t) = log\frac{\omega_t}{(1-\omega_t)} = z_t^T\gamma.$$

Also, assume that

$$f(y_t|\alpha_t) = f(y_t|\alpha_t, \boldsymbol{\alpha}^{(t-1)}) = f(y_t|\alpha_t, \boldsymbol{\alpha}^{(t-1)}, \mathbf{y}^{(t-1)}), \quad t = 1, 2, \cdots \tag{4.1.3}$$

where $\mathbf{y}^{(t)} = (y_t, y_{t-1}, \cdots, y_1)$ and $\boldsymbol{\alpha}^{(t)} = (\alpha_t, \alpha_{t-1}, \cdots, \alpha_0, \alpha_{-1}, \cdots, \alpha_{1-p})$. We call Model [4.1.2] as Poisson autoregressive hurdle model (PARH).

4.1.2 Negative binomial autoregressive hurdle model

The following model is proposed to handle temporal dependence when $f_2(\cdot)$ in Model [4.1.1] is specified as negative binomial. To proceed, assume that Y_t, x_t, z_t and α_t are defined as mentioned in the previous subsection. Conditioning on α_t, suppose Y_t is a sequence of independent counts with negative binomial hurdle

(NBH) distribution defined as follows:

$$f(y_t|\alpha_t, x_t, z_t) = \begin{cases} \omega_t, & \text{if } y_t = 0 \\ (1-\omega_t)\binom{y_t + r - 1}{y_t} \frac{p_t^r(1-p_t)^{y_t}}{(1-p_t^r)}, & \text{if } y_t > 0 \end{cases} \tag{4.1.4}$$

where

$$log\left\{\frac{r(1-p_t)}{p_t}\right\} = x_t^T\beta + \alpha_t,$$

and

$$logit(\omega_t) = z_t^T\gamma.$$

Also assume the validity of condition 4.1.3 in Model [4.1.2]. We call such model as negative binomial autoregressive hurdle model (NBARH).

4.2 Estimation

In order to estimate the parameters of PARH and NBARH models by data cloning method, we need to write them as hierarchical models. PARH model could be written as follows

$$Y_t|\alpha_t, x_t, z_t \ \sim \ \text{PH}(\lambda_t),$$

with

$$\lambda_t = \exp(x_t^T\beta + \alpha_t) \text{ and } logit(\omega_t) = z_t^T\gamma,$$

$$\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p} \sim \text{Normal}(\phi_1\alpha_{t-1} + \cdots, \phi_p\alpha_{t-p}, \sigma^2).$$

The hierarchical model for NBARH is

$$Y_t|\alpha_t, x_t, z_t, \sim \text{NBH}(\lambda_t),$$

with

$$\tfrac{r(1-p_t)}{p_t} = \exp(x_t^T\beta + \alpha_t) \text{ and } logit(\omega_t) = z_t^T\gamma,$$

$$\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p} \sim \text{Normal}(\phi_1\alpha_{t-1} + \cdots, \phi_p\alpha_{t-p}, \sigma^2).$$

The likelihood function of the models is obtained by

$$L(\theta, y) = \int \prod_{t=1}^{n} f(y_t|\alpha_t)g(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p})g_0(\alpha_0, \cdots, \alpha_{1-p})\ d\alpha,$$

where $f(y_t|\alpha_t)$ is PH density function for PARH model, and NBH density function for NBARH model, $g(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p})$ is the density function of normal distribution and $g_0(\alpha_0, \cdots, \alpha_{1-p})$ is the prior distribution for the initial conditions of the AR(p) process. $\theta = (\theta_1, \theta_2)$, where $\theta_1 = (\beta, \gamma)$ denotes the parameters of the fixed effects, and $\theta_2 = (\phi_1, \cdots, \phi_p, \sigma)$ denotes the parameters of the autoregressive latent process $\{\alpha_t\}$. For more details see Appendix A.

The posterior distribution of θ conditional on the data $y^{(K)} = (y, \cdots, y)$ is given by

$$\pi_K(\theta|y) = \frac{[L(\theta, y)]^K \pi(\theta)}{C(K, y)},$$

where $C(K, y) = \int [L(\theta, y)]^K \pi(\theta)d\theta$, is the normalizing constant.

4.3 Numerical studies

A simulation experiment is presented to check the performance of DC method when the data is simulated from PARH and NBARH models. Five hundred real-

izations were generated from each model with sample size of 500 in each realization. The simulations described below were implemented using *Jags* software and *dclone*, *rjags* and *coda* packages from R software.

4.3.1 PARH model

In this experiment we considered two cases:

1. Case 1: PARH with AR(1) latent process.

2. Case 2: PARH with AR(2) latent process.

In Case 1, the following explanatory variables were used: $x_{t1} = 1, x_{t2}$ is *uniform*(2,4) random variable, $z_{t1} = 1$ and z_{t2} is standard normal random variable, while in Case 2 we let $z_t = x_t = (1, x_{t2})$, where x_{t2} is a standard normal random variable.

The priors that we used in these simulations are: normal distribution with mean zero and variance 10^3 for fixed effects parameters, log normal distribution with mean zero and variance 1 for the inverse of the variance component. For the latent process AR(1), we used *uniform*(-0.99, 0.99) prior distribution for ϕ_1, and for ϕ_1, ϕ_2 in the latent process AR(2), we used normal prior distribution with mean zero and variance 10^4. Furthermore, we assigned normal priors distribution with mean zero and variance 1 for the initial conditions in both processes.

In each case we set the following: burn-in period of 10000, three parallel MCMC chains and then every 10th sample was kept until 1000 observations were obtained from each chain. Thus, a total of 3000 observations were generated from the joint posterior distribution of the parameters.

Tables 4.1 and 4.3 display the real values of the parameters, the empirical means, the empirical standard deviations, MSE and DC standard errors with K=3

and K=2, respectively. Both Tables show that the true value of the parameter is very close to the estimated value, and the DC standard errors and the empirical standard deviations are in very good agreement.

Tables 4.2 and 4.4 display the empirical means, the empirical standard deviations, MSE and Bayes MCMC standard errors with K=1.

Also, we present the percentage coverage of 95% confidence interval of both cases in Table 4.5.

Table 4.1: Estimation of PARH model parameters with AR(1) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
γ_0	-1.00	-1.0070	0.0154	0.1239	0.1307
γ_1	-2.00	-2.0318	0.03619	0.1878	0.1912
β_0	-1.00	-0.9705	0.0611	0.2457	0.2290
β_1	1.50	1.4907	0.0064	0.0794	0.0747
σ	0.80	0.7969	0.0011	0.0337	0.0361
ϕ_1	-0.50	-0.4922	0.0029	0.0535	0.0531

Table 4.2: Estimation of PARH model parameters with AR(1) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
γ_0	-1.00	-1.0121	0.0157	0.1248	0.1311
γ_1	-2.00	-2.0452	0.0378	0.1894	0.1920
β_0	-1.00	-0.9640	0.0668	0.2563	0.2220
β_1	1.50	1.4885	0.0070	0.0831	0.0724
σ	0.80	0.8014	0.0011	0.0339	0.0365
ϕ_1	-0.50	-0.4881	0.0030	0.0539	0.0536

Table 4.3: Estimation of PARH model parameters with AR(2) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
γ_0	-1.50	-1.5222	0.0286	0.1679	0.1734
γ_1	-3.00	-3.0610	0.0832	0.2821	0.2908
β_0	-1.00	-0.9984	0.0284	0.1688	0.1635
β_1	2.00	2.0003	0.0152	0.1236	0.1247
σ	0.90	0.9013	0.0064	0.0802	0.0793
ϕ_1	1.00	0.9950	0.0030	0.0549	0.0539
ϕ_2	-0.75	-0.7427	0.0031	0.0556	0.0499

Table 4.4: Estimation of PARH model parameters with AR(2) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
γ_0	-1.50	-1.5304	0.0295	0.1693	0.1741
γ_1	-3.00	-3.0805	0.0875	0.2849	0.2926
β_0	-1.00	-1.0089	0.0243	0.1559	0.1632
β_1	2.00	2.0040	0.0150	0.1226	0.1256
σ	0.90	0.9140	0.0064	0.0788	0.0805
ϕ_1	1.00	0.9896	0.0031	0.0550	0.0547
ϕ_2	-0.75	-0.7372	0.0029	0.0520	0.0500

Table 4.5: Percentage of coverage of a 95% confidence interval in PARH model

Parameters	AR(1) latent process	AR(2) latent process
γ_0	0.950	0.960
γ_1	0.952	0.968
β_0	0.918	0.940
β_1	0.926	0.956
σ	0.962	0.948
ϕ_1	0.952	0.934
ϕ_2	-	0.938

4.3.2 NBARH model

In this experiment we also considered two cases:

1. Case 1: NBARH with AR(1) latent process.

2. Case 2: NBARH with AR(2) latent process.

In Case 1, we used the following covariate sequences defined by $x_t^T = (1, x_{t2})$, where x_{t2} is standard normal random variable and $z_t^T = (1, z_{t2}, z_{t3})$, where $z_{t2} = cos\frac{2\pi t}{365}$ and $z_{t3} = sin\frac{2\pi t}{365}$. In Case 2, x_{t2} is $uniform$(1,2) random variable and $z_t^T = (1, z_{t2})$, where $z_{t2} = \frac{t}{500}$.

The same priors used in Section 4.3.1 were used in this experiment. Moreover, a burn-in period of 5000 samples was used with three parallel chains and then every 10th sample was kept, until 1000 observations were obtained from each chain. Hence, a total of 3000 observations were generated from the joint posterior distribution of the parameters.

The real values of the parameters, the empirical means, the empirical standard deviations, MSE and DC standard errors with K=2 are shown in Tables 4.6 and 4.8, while Tables 4.7 and 4.9 show the same summary but with K=1. Almost we have similar results for both DC method and Bayes MCMC method except for estimating ϕ_1 in AR(1) process and ϕ_1, ϕ_2 in AR(2) process, DC method gives better unbiased estimate for these parameters than Bayes MCMC method. The percentage of coverage of 95% confidence interval of both cases is presented in Table 4.10.

Table 4.6: Estimation of NBARH model parameters with AR(1) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
γ_0	-1.50	-1.5425	0.0346	0.1815	0.1721
γ_1	2.00	2.0587	0.0559	0.2296	0.2188
γ_2	1.00	1.0346	0.0408	0.1993	0.1888
β_0	-1.00	-0.9977	0.0265	0.1631	0.1577
β_1	1.50	1.5029	0.0126	0.1125	0.1141
σ	0.70	0.6542	0.0232	0.1455	0.1458
ϕ_1	-0.40	-0.3147	0.0540	0.2167	0.2863

Table 4.7: Estimation of NBARH model parameters with AR(1) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
γ_0	-1.50	-1.5535	0.0363	0.1832	0.1730
γ_1	2.00	2.0715	0.0583	0.2311	0.2196
γ_2	1.00	1.0427	0.042	0.2011	0.1896
β_0	-1.00	-1.0057	0.0268	0.1641	0.1591
β_1	1.50	1.5081	0.0128	0.1132	0.1148
σ	0.70	0.6612	0.0217	0.1425	0.1413
ϕ_1	-0.40	-0.2578	0.0565	0.1908	0.2786

Table 4.8: Estimation of NBARH model parameters with AR(2) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
γ_0	-0.50	-0.5076	0.0437	0.2091	0.1945
γ_1	-1.00	-1.0163	0.1483	0.3852	0.3581
β_0	1.00	0.9985	0.0487	0.2210	0.2163
β_1	1.50	1.5005	0.0202	0.1421	0.1374
σ	0.50	0.5026	0.0025	0.0502	0.0503
ϕ_1	1.00	0.9878	0.0110	0.1045	0.1031
ϕ_2	-0.45	-0.4469	0.0080	0.0895	0.0872

Table 4.9: Estimation of NBARH model parameters with AR(2) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
γ_0	-0.50	-0.5085	0.0440	0.2097	0.1947
γ_1	-1.00	-1.0195	0.1496	0.3867	0.3585
β_0	1.00	0.9968	0.0491	0.2217	0.2181
β_1	1.50	1.5015	0.0203	0.1426	0.1385
σ	0.50	0.5098	0.0025	0.0491	0.0503
ϕ_1	1.00	0.9741	0.0112	0.1026	0.1034
ϕ_2	-0.45	-0.4351	0.0081	0.0886	0.0879

Table 4.10: Percentage of coverage of a 95% confidence interval in NBARH model

Parameters	AR(1) latent process	AR(2) latent process
γ_0	0.942	0.936
γ_1	0.928	0.934
γ_2	0.930	-
β_0	0.960	0.954
β_1	0.948	0.948
σ	0.930	0.954
ϕ_1	0.952	0.942
ϕ_2	-	0.940

4.4 Application to asthma dataset

We revisited the dataset of daily counts of emergency department visits of asthma for age group (70-79) years in Ontario, Canada again. We began by fitting PARH model, retaining all the covariates used in ZIP model before. Namely: $x_{1t} = \frac{t}{2191}$; x_{2t} = weekend; $x_{3t} = cos\frac{2\pi t}{365}$; $x_{4t} = sin\frac{2\pi t}{365}$; $x_{5t} = cos\frac{4\pi t}{365}$ and $x_{6t} = sin\frac{4\pi t}{365}$, with AR(1) latent process and number of clones set to 1. Once we obtained our final fitted PARH model we fitted NBARH model under the same specifications as in the PARH using the same covariates, latent process and $K = 1$ clones.

In order to find an estimate for the dispersion parameter r in NBARH model, we followed the subsequent steps:

1. Find the MLE for β, γ, ϕ and σ with different values of $r, (r = 1, 2, \cdots, 9)$. (The number of clones we used here is 1).

2. Generate 100 random data samples of the latent process $\{\alpha_t\}$.

3. Calculate AIC difference for all pairs of models with different values of r.

4. Determine the estimate $\widehat{r}$ by the r value that yielded the positive AIC difference with all other values of r.

The estimate $\widehat{r} = 1$ was chosen in this analysis because it gave positive AIC difference with all other values of r.

Estimated parameters, their standard errors and Wald statistic for PARH and NBARH models are presented in Table 4.11. The diagnostic measurements for data cloning algorithm convergence and Brook-Gelman statistic, $\widehat{R}$, are shown in Table 4.12. Furthermore, trace plots of all the parameters for both models are presented in Figure 4.1 and 4.3 showing good mixing for all parameters samples with less quality of mixing for σ in NBARH model. Also, plots of posterior densities of the models parameters are shown in Figures 4.2 and 4.4.

The following priors were used for both models: $N(0, 10^3)$ for fixed effects β and γ; *log normal*(0,1) for the variance component, *Uniform*(-0.99,0.99) for the correlation component, and for the initial condition α_0, we used the prior $N(0, 1)$. For PARH model we set the following: number of chains=3; number of adaptation steps to tune the sampler=20000; number of samples to be kept after the burn-in samples are discarded=20000 and thinning rate =20. While for NBARH model we used the following: number of chains=3; burn-in period=40000; number of iterations following burn-in period=50000 and thinning rate=100. In addition, taking model 1 to be PARH model and model 2 to be NBARH model, the

difference $AIC_1 - AIC_2$ value was 4971.6 indicating that NBARH model provides better description of the data than does PARH model.

In Table 4.11, we see almost similar results for the fixed effects parameters for both models. The hurdle portion in both models exhibits significant linear decreasing trend, indicating that that for every new day, we expect $0.36\%, (1 - exp(-0.44))$, reduction in the odds of having less than five visits to emergency departments across Ontario, due to asthma in the age group (70-79) years.

Also, there is evidence of significant seasonal effects (annual and semi-annual frequencies) on the model of the zeros while the effect of the weekends were not significant on the distribution of the zeros. For the non-zero distributions, the trend is significant at 0.1 level in PARH model *(p-value=0.09)* indicating an increase in the number of daily visits to emergency departments over time, while such time trend is not significant in NBARH model. Also, there are significant annual periodic trend, while the semi-annual and weekend effects are not significant in both models. The temporal autocorrelation among the non-zero observations is also statistically significant with larger correlation value in NBARH model. Similarly, there are significant variance components in both models indicating heterogeneity in the non-zero observations.

Table 4.11: Estimates and their standard errors from analysis of asthma data by PARH and NBARH parameter driven models

Parameter	PARH model			NBARH model		
	MLE	DC SE	z-value	MLE	DC SE	z-value
β_0	0.8139	0.0745	10.92**	0.5801	0.1769	5.37**
β_1	0.2063	0.1234	1.67	0.2426	0.1706	1.42
β_2	-0.0341	0.0438	-0.78	-0.0503	0.0761	-0.66
β_3	0.1728	0.0516	3.35**	0.2359	0.0727	3.25**
β_4	0.2179	0.0493	4.42**	0.2563	0.0686	3.74**
β_5	-0.0291	0.0479	-0.61	-0.0099	0.0652	-0.15
β_6	-0.0549	0.0506	-1.09	-0.0802	0.0674	-1.19
γ_0	-0.1299	0.0931	-1.40	-0.1285	0.0930	-1.38
γ_1	-0.4432	0.1561	-2.84**	-0.4447	0.1557	-2.86**
γ_2	-0.0978	0.0988	-0.99	-0.0991	0.0983	-1.01
γ_3	-0.1898	0.0621	-3.06**	-0.1891	0.0619	-3.06**
γ_4	-0.5364	0.0647	-8.29**	-0.5368	0.0645	-8.32**
γ_5	0.1491	0.0632	2.36*	0.1486	0.0630	2.36*
γ_6	0.0335	0.0625	0.54	0.0335	0.0632	0.53
ϕ_1	0.7624	0.0535	14.26**	0.9338	0.0307	30.45**
σ	0.2987	0.0397	7.53**	0.0963	0.0242	3.99**

* indicates significant at 0.05 level ** indicates significant at 0.01 level

Table 4.12: Estimability diagnostics

Model	Number of clones	lambda.max	ms error	r^2	$\widehat{R}$
PARH	1	0.0311	0.0426	0.0004	1.0085
NBARH	1	0.0384	0.0857	0.0204	1.0315

Figure 4.1: The trace plots of PARH model parameters

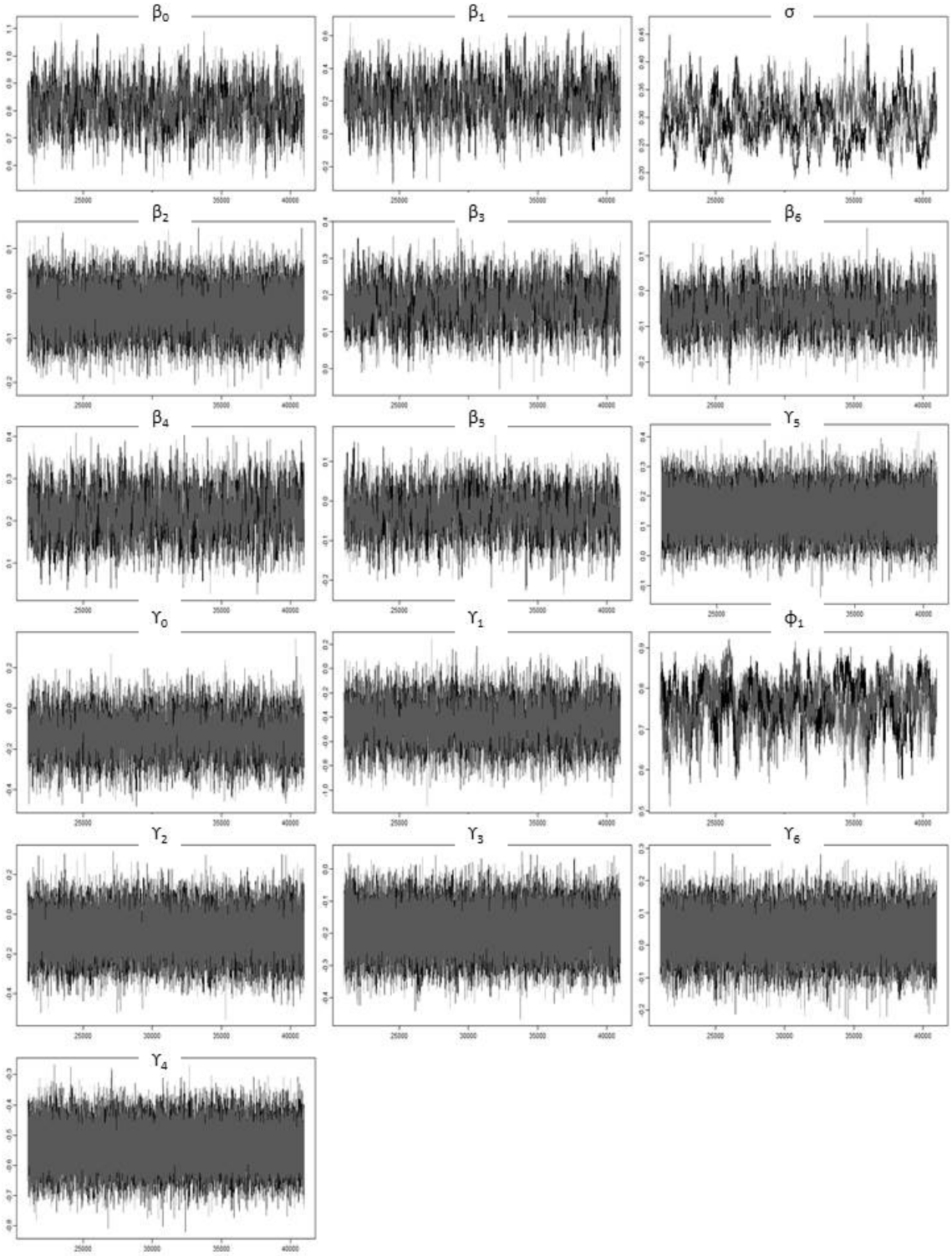


Figure 4.2: The posterior densities of PARH model parameters

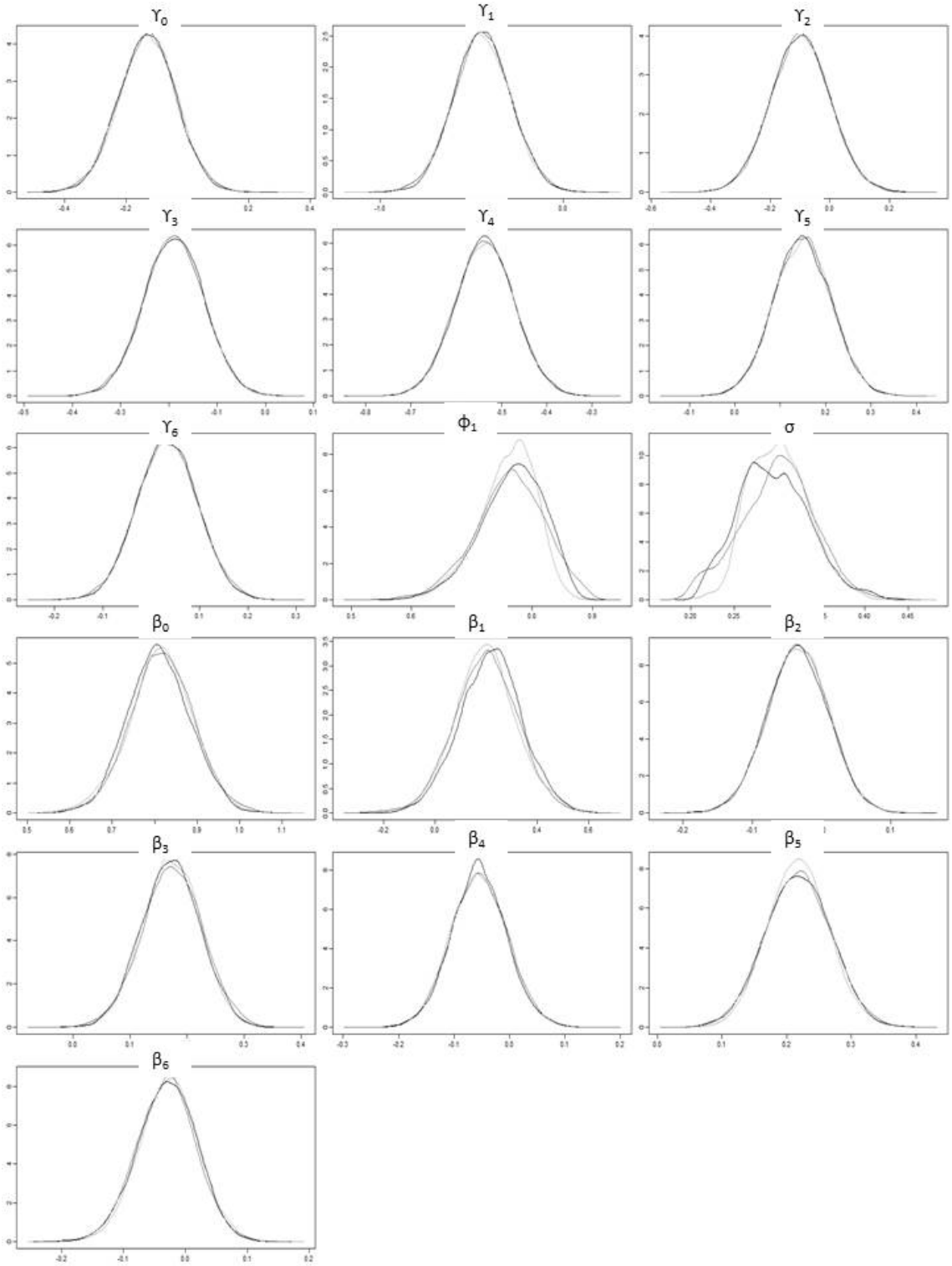


Figure 4.3: The trace plots of NBARH model parameters

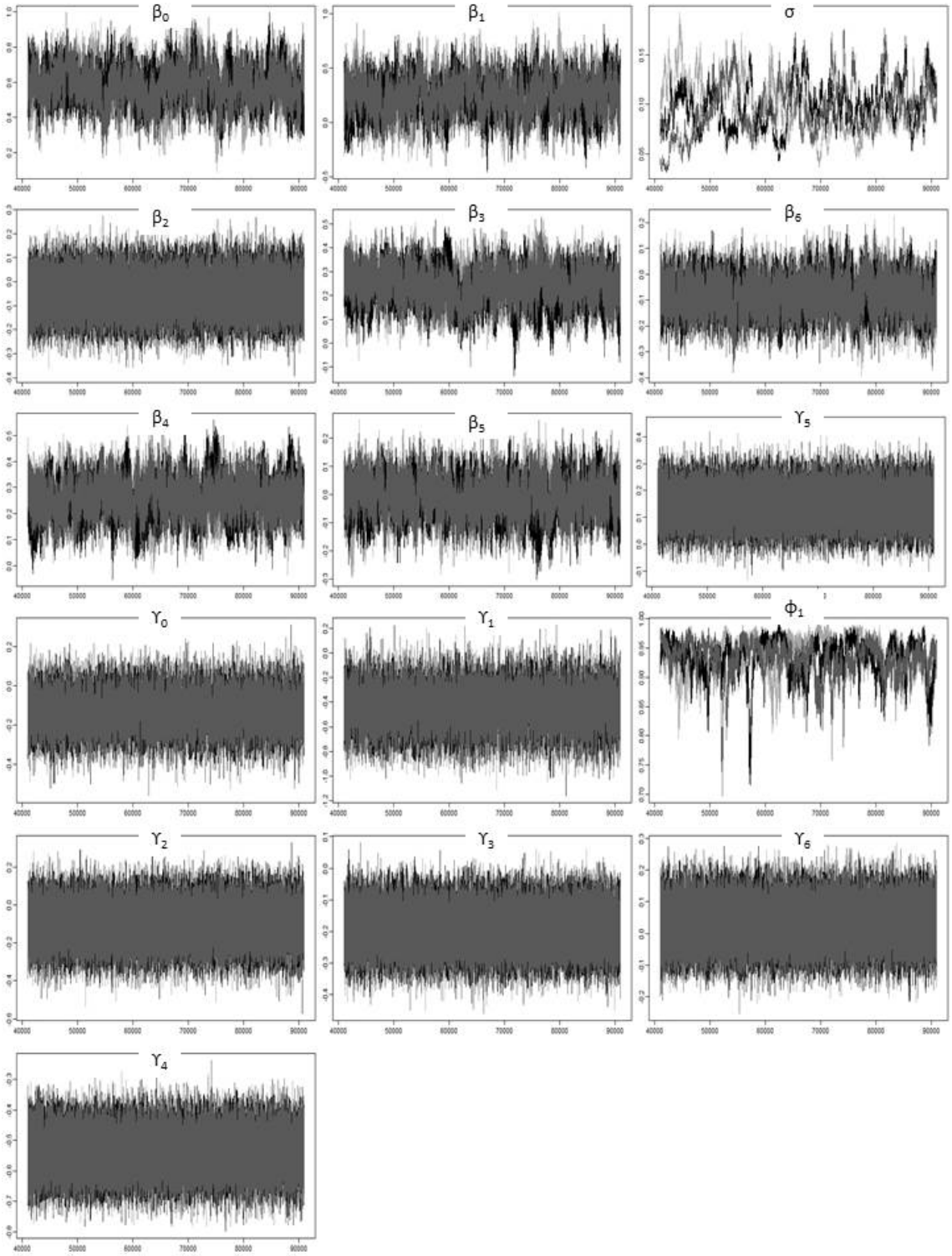
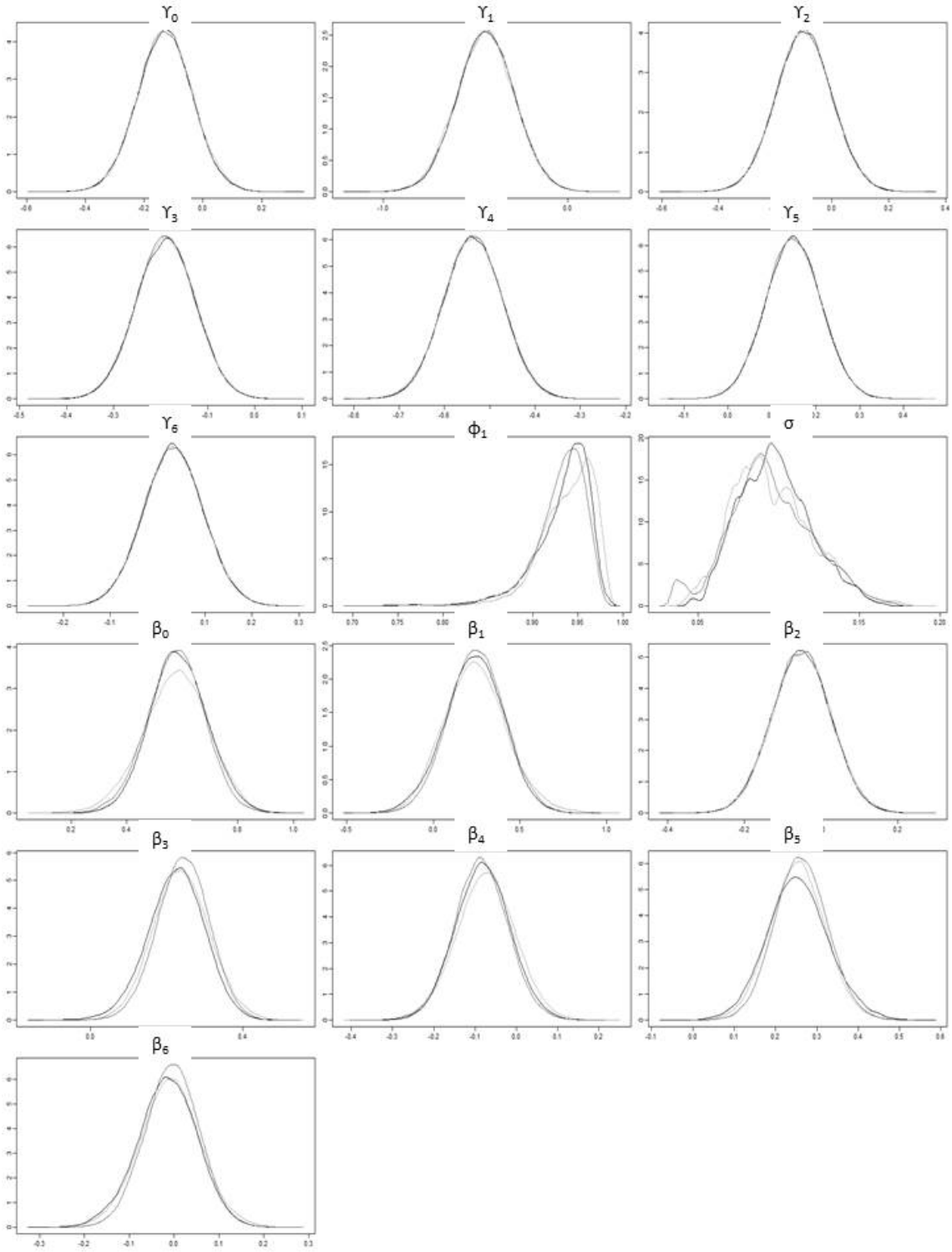


Figure 4.4: The posterior densities of NBARH model parameters



Chapter 5

Parameter-driven Bivariate Poisson model

The organization of this chapter is as follows: in Section 1 we present two parameter-driven bivariate Poisson models and derive their properties. The estimation procedure for the unknown parameters is again based on the DC algorithm and is discussed in Section 2. A simulation study and real data applications are provided in Sections 3 and 4.

5.1 Bivariate Poisson models

Suppose that W_1, W_2 and W_3 are three independent random variables following Poisson distributions with parameters λ_1, λ_2 and λ_3. Define $Y = W_1 + W_3$ and $Z = W_2 + W_3$. Then the joint probability mass function of Y and Z can be derived as follows

$$P(Y = y, Z = z) = P(W_1 + W_3 = y, W_2 + W_3 = z)$$

$$= \sum_{i=0}^{min(y,z)} P(W_1 = y-i, W_2 = z-i, W_3 = i)$$
$$= \sum_{i=0}^{min(y,z)} P(W_1 = y-i)P(W_2 = z-i)P(W_3 = i).$$

Of course, we get the last step using the independence of W_1, W_2 and W_3. Now, the joint probability mass function can be written as

$$P(Y = y, Z = z) = \sum_{i=0}^{min(y,z)} \frac{e^{-\lambda_1}\lambda_1^{y-i}}{(y-i)!} \frac{e^{-\lambda_2}\lambda_2^{z-i}}{(z-i)!} \frac{e^{-\lambda_3}\lambda_3^{i}}{i!}$$
$$= e^{-(\lambda_1+\lambda_2+\lambda_3)} \frac{\lambda_1^y}{y!}\frac{\lambda_2^z}{z!} \sum_{i=0}^{min(y,z)} \frac{\lambda_3^i}{i!} \frac{y!\ \lambda_1^{-i}}{(y-i)!} \frac{z!\ \lambda_2^{-i}}{(z-i)!}.$$

Hence, the bivariate Poisson distribution of Y and Z is given by

$$P(Y = y, Z = z) = e^{-(\lambda_1+\lambda_2+\lambda_3)} \frac{\lambda_1^y}{y!}\frac{\lambda_2^z}{z!} \sum_{i=0}^{min(y,z)} \binom{y}{i}\binom{z}{i} i! \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^i, \qquad (5.1.1)$$

where $y, z = 0, 1, 2, \cdots$.

Clearly, Y and Z marginally have Poisson distributions with means $\lambda_1 + \lambda_3$ and $\lambda_2 + \lambda_3$, respectively. Moreover, the covariance between Y and Z becomes

$$Cov(Y, Z) = Cov(W_1 + W_3, W_2 + W_3) = Var(W_3) = \lambda_3,$$

and thus the correlation equals

$$Corr(Y, Z) = \frac{\lambda_3}{\sqrt{(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)}}.$$

We can see that the correlation is always positive as both λ_3 and the dominator exceed zero.

The bivariate Poisson regression model arises if we assume that the parameters depend on some explanatory variables, in other words:

$$ln\lambda_i = x_i^T \beta_i = \sum_{j=1}^{p_i} \beta_{ij} x_{ij}, \quad i = 1, 2, 3$$

where the vectors x_i and β_i have dimension $p_i \times 1$. Therefore, the set of covariates and their number may be different for different λ_i's.

Suppose that $\{(Y_t, Z_t) : t = 1, 2, \cdots, n\}$ is a bivariate time series of observed counts. Then we expect some dependence between successive observations, to accommodate this kind of dependence we propose two bivariate Poisson models: the first one with one latent process added to the cross-correlation parameter λ_3, to introduce equal correlation functions in the two processes $\{Y_t\}$ and $\{Z_t\}$ and also to accommodate cross-correlation between them. The second one with two latent processes added to the parameters λ_1 and λ_2, to propose different correlation functions in the two processes $\{Y_t\}$ and $\{Z_t\}$. In addition, both models accommodate over-dispersion of the data.

5.1.1 Bivariate Poisson with one latent process

To handle the correlation and cross-correlation between successive bivariate observations, the following parameter-driven bivariate Poisson model with one latent process (BP1) is proposed. To illustrate, consider a stationary autoregressive process of order *p, AR(p)*, such that

$$\alpha_t = \phi\alpha_{t-1} + \phi_2\alpha_{t-2} + \cdots + \phi_p\alpha_{t-p} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a normal random process with mean zero and variance σ^2. Conditioning on α_t, suppose (y_t, z_t) is a sequence of independent counts with bivariate Poisson distribution defined as follows:

$$f(y_t, z_t|\alpha_t) = e^{-(\lambda_{1t}+\lambda_{2t}+\lambda_{3t})}\frac{\lambda_{1t}^{y_t}}{y_t!}\frac{\lambda_{2t}^{z_t}}{z_t!}\sum_{i=0}^{min(y_t,z_t)}\binom{y_t}{i}\binom{z_t}{i}i!\left(\frac{\lambda_{3t}}{\lambda_{1t}\lambda_{2t}}\right)^i, \qquad (5.1.2)$$

where $y_t, z_t = 0, 1, 2, \cdots$ and $t = 1, 2, \cdots, n$. The parameters $\lambda_{1t}, \lambda_{2t}$ and λ_{3t} satisfy

$$ln\lambda_{3t} = x_{3t}^T\beta_3 + \alpha_t = \sum_{j=1}^{p_3}\beta_{3j}x_{3jt} + \alpha_t,$$

and

$$ln\lambda_{it} = x_{it}^T\beta_i = \sum_{j=1}^{p_i}\beta_{ij}x_{ijt}, \quad (i = 1, 2),$$

where the vectors x_{it} and β_i $(i = 1, 2, 3)$ have dimension $p_i \times 1$. Also, assume that

$$f(y_t, z_t|\alpha_t) = f(y_t, z_t|\alpha_t, \boldsymbol{\alpha}^{(t-1)}) = f(y_t, z_t|\alpha_t, \boldsymbol{\alpha}^{(t-1)}, \mathbf{y}^{(t-1)}, \mathbf{z}^{(t-1)}), \quad t = 1, 2, \cdots \qquad (5.1.3)$$

where $(\mathbf{y}^{(t)}, \mathbf{z}^{(t)}) = ((y_t, z_t), (y_{t-1}, z_{t-1}), \cdots, (y_1, z_1))$ and
$\boldsymbol{\alpha}^{(t)} = (\alpha_t, \alpha_{t-1}, \cdots, \alpha_0, \alpha_{-1}, \cdots, \alpha_{1-p})$.

The marginal moments of the observed bivariate process $\{(Y_t, Z_t)\}$ are given as follows:

$$EY_t = E(E(Y_t|\alpha_t)) = E(\lambda_{1t} + \lambda_{3t}) = e^{x_{1t}^T\beta_1} + e^{x_3t^T\beta_3}Ee^{\alpha_t} = e^{x_{1t}^T\beta_1} + \mu_\alpha e^{x_{3t}^T\beta_3},$$

where $\mu_\alpha = Ee^{\alpha_t}$, and

$$\sigma_{y_t}^2 = Var(Y_t) = E(Var(Y_t|\alpha_t)) + Var(E(Y_t|\alpha_t))$$

$$
\begin{aligned}
&= E(\lambda_{1t} + \lambda_{3t}) + Var(\lambda_{1t} + \lambda_{3t}) \\
&= e^{x_{1t}^T \beta_1} + e^{x_{3t}^T \beta_3} E e^{\alpha_t} + e^{2x_{3t}^T \beta_3} Var(e^{\alpha_t}) \\
&= e^{x_{1t}^T \beta_1} + (\mu_\alpha + \sigma_\alpha^2 e^{x_{3t}^T \beta_3}) e^{x_{3t}^T \beta_3},
\end{aligned}
$$

where $\sigma_\alpha^2 = Var(e^{\alpha_t})$. Similarly, the mean and the variance of the process $\{Z_t\}$ are

$$
EZ_t = e^{x_{2t}^T \beta_2} + \mu_\alpha e^{x_{3t}^T \beta_3},
$$

and

$$
\sigma_{z_t}^2 = Var(Z_t) = e^{x_{2t}^T \beta_2} + (\mu_\alpha + \sigma_\alpha^2 e^{x_{3t}^T \beta_3}) e^{x_{3t}^T \beta_3}.
$$

The autocovariance function of the observed process $\{Y_t\}$ is given by

$$
\begin{aligned}
\gamma_{y_t}(h) &= Cov(E(Y_t|\boldsymbol{\alpha}^{(t+h)}), E(Y_{t+h}|\boldsymbol{\alpha}^{(t+h)})) + E(Cov(Y_t|\boldsymbol{\alpha}^{(t+h)}, Y_{t+h}|\boldsymbol{\alpha}^{(t+h)})) \\
&= Cov(E(Y_t|\alpha_t), E(Y_{t+h}|\alpha_{t+h})) + E(Cov(Y_t|\alpha_t, Y_{t+h}|\alpha_{t+h})) \\
&= Cov(\lambda_{1t} + \lambda_{3t}, \lambda_{1(t+h)} + \lambda_{3(t+h)}) \\
&= Cov(e^{x_{3t}^T \beta_3 + \alpha_t}, e^{x_{3(t+h)}^T \beta_3 + \alpha_{t+h}}) \\
&= e^{(x_{3t} + x_{3(t+h)})^T \beta_3} Cov(e^{\alpha_t}, e^{\alpha_{t+h}}) \\
&= e^{(x_{3t} + x_{3(t+h)})^T \beta_3} \gamma_\alpha(h),
\end{aligned}
$$

where $\gamma_\alpha(h)$ is the autocovariance function of the latent process $\{e^{\alpha_t}\}$ and $h \neq 0$. In the above derivation, we have used the fact that the conditional distribution of the observed process Y_t given the future of the latent processes is same as if only the current states of the latent processes were given (for details see Proposition B.1. in Appendix). Clearly, the correlation between Y_t and Y_{t+h} could be positive or negative depending on the sign of $\gamma_\alpha(h)$, and the process $\{Y_t\}$ is not stationary

time series since $\gamma_{y_t}(h)$ depends on t. In addition, following the previous steps, we will get the same autocovariance function for the observed process $\{Z_t\}$.

Furthermore, the cross-covariance function between Y_t and Z_{t+h} when $h = 0$, can be derived as follows

$$\begin{aligned}
\gamma_{y_t z_t}(0) =& Cov(E(Y_t|\boldsymbol{\alpha}^{(t+h)}), E(Z_t|\boldsymbol{\alpha}^{(t+h)})) + E(Cov(Y_t|\boldsymbol{\alpha}^{(t+h)}, Z_t|\boldsymbol{\alpha}^{(t+h)})) \\
=& Cov(\lambda_{1t} + \lambda_{3t}, \lambda_{2t} + \lambda_{3t}) + E(\lambda_{3t}) \\
=& Cov(e^{x_{3t}^T\beta_3+\alpha_t}, e^{x_{3t}^T\beta_3+\alpha_t}) + E(e^{x_{3t}^T\beta_3+\alpha_t}) \\
=& e^{2x_{3t}^T\beta_3} Var(e^{\alpha_t}) + e^{x_{3t}^T\beta_3} E(e^{\alpha_t}) \\
=& (\mu_\alpha + \sigma_\alpha^2 e^{x_{3t}^T\beta_3}) e^{x_{3t}^T\beta_3},
\end{aligned}$$

showing that the cross-correlation between Y_t and Z_t is always positive. When $h \neq 0$, the cross-covariance function is

$$\begin{aligned}
\gamma_{y_t z_t}(h) = Cov(Y_t, Z_{t+h}) =& Cov(E(Y_t|\alpha_t), E(Z_{t+h}|\alpha_{t+h})) + E(Cov(Y_t|\alpha_t, Z_{t+h}|\alpha_{t+h})) \\
=& Cov(\lambda_{1t} + \lambda_{3t}, \lambda_{2(t+h)} + \lambda_{3(t+h)}) \\
=& Cov(e^{x_{3t}^T\beta_3+\alpha_t}, e^{x_{3(t+h)}^T\beta_3+\alpha_{t+h}}) \\
=& e^{(x_{3t}+x_{3(t+h)})^T\beta_3} Cov(e^{\alpha_t}, e^{\alpha_{t+h}}) \\
=& e^{(x_{3t}+x_{3(t+h)})^T\beta_3} \gamma_\alpha(h),
\end{aligned}$$

which is the same autocovariance function for the processes $\{Y_t\}$ and $\{Z_t\}$.

As a special case, assume that $\alpha_t = \phi\alpha_{t-1} + \epsilon_t$ is an $AR(1)$ process, the marginal moments of (Y_t, Z_t) are obtained as follows

$$E(Y_t) = e^{x_{1t}^T\beta_1} + e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{3t}^T\beta_3} \quad \text{and} \quad E(Z_t) = e^{x_{2t}^T\beta_2} + e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{3t}^T\beta_3},$$

$$\begin{aligned}\sigma^2_{y_t} = Var(Y_t) =& e^{x_{1t}^T\beta_1} + e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{3t}^T\beta_3} + e^{\frac{\sigma^2}{(1-\phi^2)}+2x_{3t}^T\beta_3}\left(e^{\frac{\sigma^2}{(1-\phi^2)}} - 1\right) \\ =& e^{x_{1t}^T\beta_1} + \left[1 + e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{3t}^T\beta_3}\left(e^{\frac{\sigma^2}{(1-\phi^2)}} - 1\right)\right] e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{3t}^T\beta_3},\end{aligned}$$

similarly,

$$\sigma^2_{z_t} = Var(Z_t) = e^{x_{2t}^T\beta_2} + \left[1 + e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{3t}^T\beta_3}\left(e^{\frac{\sigma^2}{(1-\phi^2)}} - 1\right)\right] e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{3t}^T\beta_3}.$$

The autocovariance function of the process $\{Y_t\}$ is

$$\gamma_{y_t}(h) = e^{(x_{3t}+x_{3(t+h)})^T\beta_3+\frac{\sigma^2}{(1-\phi^2)}}\left(e^{\frac{\sigma^2\phi^h}{(1-\phi^2)}} - 1\right),$$

which is also the autocovariance function for $\{Z_t\}$ and is the cross-covariance function between Y_t and $Z_{t+h}, h \neq 0$. Finally, the cross-covariance function for Y_t and Z_t is

$$\begin{aligned}\gamma_{y_t z_t}(0) &= \left[e^{\frac{\sigma^2}{2(1-\phi^2)}} + e^{\frac{\sigma^2}{(1-\phi^2)}+x_{3t}^T\beta_3}\left(e^{\frac{\sigma^2}{(1-\phi^2)}} - 1\right)\right] e^{x_{3t}^T\beta_3} \\ &= \left[1 + e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{3t}^T\beta_3}\left(e^{\frac{\sigma^2}{(1-\phi^2)}} - 1\right)\right] e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{3t}^T\beta_3}.\end{aligned}$$

5.1.2 Bivariate Poisson with two latent processes

In this section, we introduce another parameter-driven bivariate Poisson model with two latent processes (BP2) to accommodate different correlations in the observed processes $\{Y_t\}$ and $\{Z_t\}$. To be precise, consider two stationary autoregressive processes of order p and q, respectively, such that

$$\alpha_t = \phi_1\alpha_{t-1} + \phi_2\alpha_{t-2} + \cdots + \phi_p\alpha_{t-p} + \epsilon_t,$$

and

$$\tilde{\alpha}_t = \tilde{\phi}_1\tilde{\alpha}_{t-1} + \tilde{\phi}_2\tilde{\alpha}_{t-2} + \cdots + \tilde{\phi}_p\tilde{\alpha}_{t-q} + \tilde{\epsilon}_t,$$

where $\{\epsilon_t\}$ and $\{\tilde{\epsilon}_t\}$ are independent normal random processes with mean zero and variance σ^2 and $\tilde{\sigma}^2$, respectively. Suppose that (y_t, z_t) conditioning on α_t and $\tilde{\alpha}_t$ from a bivariate sequence of independent counts with bivariate Poisson distribution defined by:

$$f(y_t, z_t|\alpha_t, \tilde{\alpha}_t) = e^{-(\lambda_{1t}+\lambda_{2t}+\lambda_{3t})}\frac{\lambda_{1t}^{y_t}}{y_t!}\frac{\lambda_{2t}^{z_t}}{z_t!}\sum_{i=0}^{min(y_t,z_t)}\binom{y_t}{i}\binom{z_t}{i}i!\left(\frac{\lambda_{3t}}{\lambda_{1t}\lambda_{2t}}\right)^i, \quad (5.1.4)$$

where $y_t, z_t = 0, 1, 2, \cdots$ and $t = 1, 2, \cdots, n$. The parameters $\lambda_{1t}, \lambda_{2t}$ and λ_{3t} satisfy

$$ln\lambda_{1t} = x_{1t}^T\beta_1 + \alpha_t = \sum_{j=1}^{p_1}\beta_{1j}x_{1jt} + \alpha_t, \quad ln\lambda_{2t} = x_t^T\beta_2 + \tilde{\alpha}_t = \sum_{j=1}^{p_2}\beta_{2j}x_{2jt} + \tilde{\alpha}_t$$

and

$$ln\lambda_{3t} = x_{3t}^T\beta_3 = \sum_{j=1}^{p_3}\beta_{3j}x_{3jt},$$

where the vectors x_{it} and β_i $(i = 1, 2, 3)$ have dimension $p_i \times 1$. Also, consider the following assumptions:

1. $f(y_t, z_t|\alpha_t, \tilde{\alpha}_t) = f(y_t, z_t|\boldsymbol{\alpha}^{(t)}, \tilde{\boldsymbol{\alpha}}^{(t)}) = f(y_t, z_t|\boldsymbol{\alpha}^{(t)}, \tilde{\boldsymbol{\alpha}}^{(t)}, \mathbf{y}^{(t-1)}, \mathbf{z}^{(t-1)})$
2. $f(y_t|\alpha_t) = f(y_t|\boldsymbol{\alpha}^{(t)}, \tilde{\boldsymbol{\alpha}}^{(t)})$
3. $f(z_t|\tilde{\alpha}_t) = f(z_t|\boldsymbol{\alpha}^{(t)}, \tilde{\boldsymbol{\alpha}}^{(t)})$

The marginal moments of the observed bivariate process $\{(Y_t, Z_t)\}$ are given as follows:

$$EY_t = E(E(Y_t|\alpha_t)) = E(\lambda_{1t} + \lambda_{3t}) = e^{x_{1t}^T\beta_1}Ee^{\alpha_t} + e^{x_3t^T\beta_3} = e^{x_{3t}^T\beta_3} + \mu_\alpha e^{x_{1t}^T\beta_1},$$

where $\mu_\alpha = Ee^{\alpha_t}$. Similarly, the marginal mean of Z_t is

$$EZ_t = e^{x_{3t}^T\beta_3} + \mu_{\tilde{\alpha}} e^{x_{2t}^T\beta_2}, \quad \text{where} \quad \mu_{\tilde{\alpha}} = Ee^{\tilde{\alpha}}.$$

The variance of the process $\{Y_t\}$ is

$$\begin{aligned}
\sigma_{y_t}^2 = Var(Y_t) =& E(Var(Y_t|\alpha_t)) + Var(E(Y_t|\alpha_t)) \\
=& E(\lambda_{1t} + \lambda_{3t}) + Var(\lambda_{1t} + \lambda_{3t}) \\
=& e^{x_{1t}^T\beta_1} Ee^{\alpha_t} + e^{x_{3t}^T\beta_3} + e^{2x_{1t}^T\beta_1} Var(e^{\alpha_t}) \\
=& e^{x_{3t}^T\beta_3} + (\mu_\alpha + \sigma_\alpha^2 e^{x_{1t}^T\beta_1}) e^{x_{1t}^T\beta_1},
\end{aligned}$$

where $\sigma_\alpha^2 = Var(e^{\alpha_t})$. Analagously, the marginal variance of Z_t is

$$\sigma_{z_t}^2 = Var(Z_t) = e^{x_{3t}^T\beta_3} + (\mu_{\tilde{\alpha}} + \sigma_{\tilde{\alpha}}^2 e^{x_{2t}^T\beta_2}) e^{x_{2t}^T\beta_2}, \quad \text{where} \quad \sigma_{\tilde{\alpha}}^2 = Var(e^{\tilde{\alpha}}).$$

The autocovariance function of the process $\{Y_t\}$ is obtained as follows

$$\begin{aligned}
\gamma_{y_t}(h) =& Cov(E(Y_t|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)}), E(Y_{t+h}|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})) + \\
& E(Cov(Y_t|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)}, Y_{t+h}|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})) \\
=& Cov(E(Y_t|\alpha_t), E(Y_{t+h}|\alpha_{t+h})) + E(Cov(Y_t|\alpha_t, Y_{t+h}|\alpha_{t+h})) \\
=& Cov(\lambda_{1t} + \lambda_{3t}, \lambda_{1(t+h)} + \lambda_{3(t+h)}) \\
=& Cov(e^{x_{1t}^T\beta_1 + \alpha_t}, e^{x_{1(t+h)}^T\beta_1 + \alpha_{t+h}}) \\
=& e^{(x_{1t} + x_{1(t+h)})^T\beta_1} Cov(e^{\alpha_t}, e^{\alpha_{t+h}}) \\
=& e^{(x_{1t} + x_{1(t+h)})^T\beta_1} \gamma_\alpha(h),
\end{aligned}$$

where $\gamma_\alpha(h)$ is the autocovariance function of the latent process $\{e^{\alpha_t}\}$. In the above

derivation we have used the fact that $Y_t|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)}$ has the same distribution as $Y_t|\alpha_t$ proven in Proposition B.2. in Appendix B. Similarly, the autocovariance function of $\{Z_t\}$ is given by

$$\gamma_{z_t}(h) = Cov(Z_t, Z_{t+h}) = e^{(x_{2t}+x_{2(t+h)})^T\beta_2}\gamma_{\tilde{\alpha}}(h),$$

where $\gamma_{\tilde{\alpha}}(h)$ is the autocovariance function of the latent process $\{e^{\tilde{\alpha}}\}$. Notice that we have different autocovariance functions for the two processes $\{Y_t\}$ and $\{Z_t\}$. Moreover, the cross-covariance function between Y_t and Z_t, when $h = 0$, is derived as follows

$$\begin{aligned}
\gamma_{y_t z_t}(0) =& Cov(E(Y_t|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)}), E(Z_t|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})) + \\
& E(Cov(Y_t|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)}, Z_t|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})) \\
=& Cov(E(Y_t|\alpha_t), E(Z_t|\tilde{\alpha}_t)) + E(Cov(Y_t|\alpha_t, Z_t|\tilde{\alpha}_t)) \\
=& Cov(\lambda_{1t} + \lambda_{3t}, \lambda_{2t} + \lambda_{3t}) + E(\lambda_{3t}) \\
=& Cov(e^{x_{1t}^T\beta_1+\alpha_t}, e^{x_{2t}^T\beta_2+\tilde{\alpha}_t}) + e^{x_{3t}^T\beta_3} \\
=& e^{x_{1t}^T\beta_1+x_{2t}^T\beta_2}Cov(e^{\alpha_t}, e^{\tilde{\alpha}_t}) + e^{x_{3t}^T\beta_3} \\
=& e^{x_{3t}^T\beta_3}.
\end{aligned}$$

When $h \neq 0$,

$$\begin{aligned}
\gamma_{y_t z_t}(h) =& Cov(E(Y_t|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)}), E(Z_{t+h}|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})) + \\
& E(Cov(Y_t|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)}, Z_{t+h}|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})) \\
=& Cov(E(Y_t|\alpha_t), E(Z_{t+h}|\tilde{\alpha}_{t+h})) + E(Cov(Y_t|\alpha_t, Z_{t+h}|\tilde{\alpha}_{t+h})) \\
=& Cov(\lambda_{1t} + \lambda_{3t}, \lambda_{2t} + \lambda_{3t})
\end{aligned}$$

$$
\begin{aligned}
&= Cov(e^{x_{1t}^T\beta_1+\alpha_t}, e^{x_{2(t+h)}^T\beta_2+\tilde{\alpha}_{t+h}}) \\
&= e^{x_{1t}^T\beta_1+x_{2(t+h)}^T\beta_2} Cov(e^{\alpha_t}, e^{\tilde{\alpha}_{t+h}}) \\
&= 0,
\end{aligned}
$$

which means that there is no cross-correlation betweeen the random variables Y_t and Z_{t+h} except when $h = 0$.

Consider the special case when $\alpha_t = \phi\alpha_{t-1} + \epsilon_t$ and $\tilde{\alpha}_t = \tilde{\phi}\tilde{\alpha}_{t-1} + \tilde{\epsilon}_t$. The marginal moments of the process $\{(Y_t, Z_t)\}$ are given as follows

$$
E(Y_t) = e^{x_{3t}^T\beta_3} + e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{1t}^T\beta_1} \quad \text{and} \quad E(Z_t) = e^{x_{3t}^T\beta_3} + e^{\frac{\tilde{\sigma}^2}{2(1-\tilde{\phi}^2)}+x_{2t}^T\beta_2},
$$

$$
\sigma_{y_t}^2 = Var(Y_t) = e^{x_{3t}^T\beta_3} + \left[1 + e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{1t}^T\beta_1}\left(e^{\frac{\sigma^2}{(1-\phi^2)}} - 1\right)\right] e^{\frac{\sigma^2}{2(1-\phi^2)}+x_{1t}^T\beta_1},
$$

$$
\sigma_{z_t}^2 = Var(Z_t) = e^{x_{3t}^T\beta_3} + \left[1 + e^{\frac{\tilde{\sigma}^2}{2(1-\tilde{\phi}^2)}+x_{2t}^T\beta_2}\left(e^{\frac{\tilde{\sigma}^2}{(1-\tilde{\phi}^2)}} - 1\right)\right] e^{\frac{\tilde{\sigma}^2}{2(1-\tilde{\phi}^2)}+x_{2t}^T\beta_2}.
$$

The autocovariance function of the two processes $\{Y_t\}$ and $\{Z_t\}$ are given, respectively as follows

$$
\gamma_{y_t}(h) = e^{(x_{1t}+x_{1(t+h)})^T\beta_1+\frac{\sigma^2}{(1-\phi^2)}}\left(e^{\frac{\sigma^2\phi^h}{(1-\phi^2)}} - 1\right),
$$

$$
\gamma_{z_t}(h) = e^{(x_{2t}+x_{2(t+h)})^T\beta_2+\frac{\tilde{\sigma}^2}{(1-\tilde{\phi}^2)}}\left(e^{\frac{\tilde{\sigma}^2\tilde{\phi}^h}{(1-\tilde{\phi}^2)}} - 1\right),
$$

and the cross-covariance function is

$$
\gamma_{y_t z_t}(h) = \begin{cases} e^{x_{3t}^T\beta_3}, & h = 0 \\ 0, & h \neq 0 \end{cases}
$$

5.2 Estimation

We can think of the bivariate Poisson regression model as a hierarchical model that defines first the probability distribution function of W_3, which is Poisson with parameter λ_3, and then defines the joint probability mass function of $Y, Z|W_3$. To illustrate, we note that

$$\begin{aligned} P(Y=y, Z=z|W_3=w_3) =& P(W_1+W_3=y, W_2+W_3=z|W_3=w_3) \\ =& P(W_1=y-w_3, W_2=z-w_3) \\ =& P(W_1=y-w_3)P(W_2=z-w_2) \\ =& \frac{e^{-\lambda_1}\lambda_1^{y-w_3}}{(y-w_3)!}\frac{e^{-\lambda_2}\lambda_2^{z-w_3}}{(z-w_3)!}, \end{aligned}$$

which is the product of the univariate probability functions of W_1 and W_2. In the third step of the above derivation, we have made use of the fact that W_1 and W_2 are independent of each other. Now, the likelihood function can be written as

$$L(\theta, y, z) = \sum_{w_3=0}^{min(y,z)} P(y, z|w_3)P(w_3).$$

Notice that $y, z \geq w_3$, which implies $0 \leq w_3 \leq min(y, z)$, where the non-negativity constraint is due to the fact that the support of a Poisson distribution is the set of non-negative integers.

5.2.1 Parameter estimation of BP1 model

The BP1 model [5.1.2] can be written hierarchy as follows:

$$Y_t, Z_t|W_{3t}, \alpha_t \sim \prod_{i=1}^{2} \text{Poisson } (\lambda_i), \quad \text{with } \ ln\lambda_i = x_{it}^T\beta_i, (i=1,2),$$

$$W_{3t}|\alpha_t \sim \text{Poisson } (\lambda_3), \quad \text{with } \ ln\lambda_3 = x_{3t}^T\beta_3 + \alpha_t,$$

$$\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p} \sim \text{Normal } (\phi_1\alpha_{t-1} + \cdots + \phi_p\alpha_{t-p}, \sigma^2),$$

subject to the restriction $0 \leq w_{3t} \leq min(y_t, z_t), t = 1, 2, \cdots, n$. The likelihood function $L(\theta, y, z)$ of this model is obtained by

$$\int \sum_{w_{3t=0}}^{min(y_t,z_t)} \prod_{t=1}^{n} f(y_t, z_t|w_{3t}, \alpha_t)h(w_{3t}|\alpha_t)g(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p})g_0(\alpha_0, \cdots, \alpha_{1-p})d\alpha,$$

where $\theta = (\theta_1, \theta_2), \theta_1 = (\beta_1, \beta_2, \beta_3)$ denotes the parameters of the fixed effects and $\theta_2 = (\phi_1, \cdots, \phi_p, \sigma)$ denotes the parameters of the autoregressive latent process $\{\alpha_t\}$. $\alpha_0, \cdots, \alpha_{1-p}$ are the initial conditions of the AR(p) process. For more details see Appendix A.

The posterior distribution of θ conditional on the cloned data $(y, z)^{(K)} = ((y, z), \cdots, (y, z))$ is given by

$$\pi_K(\theta|y, z) = \frac{[L(\theta, y, z)]^K \pi(\theta)}{C(K, y, z)},$$

where $\pi(\theta)$ is the prior distribution of the parameters, and

$$C(K, y, z) = \int [L(\theta, y, z)]^K \pi(\theta) d\theta,$$

is the normalizing constant.

5.2.2 Parameter estimation of BP2 model

The BP2 model [5.1.4] can be written hierarchically as follows:

$$Y_t, Z_t | W_{3t}, \alpha_t, \tilde{\alpha}_t \sim \prod_{i=1}^{2} \text{Poisson } (\lambda_i),$$

$$\text{with } \; ln\lambda_1 = x_{1t}^T \beta_1 + \alpha_t \; \text{ and } \; ln\lambda_2 = x_{2t}^T \beta_2 + \tilde{\alpha}_t,$$

$$W_{3t} \sim \text{Poisson } (\lambda_3), \; \text{ with } \; ln\lambda_3 = x_{3t}^T \beta_3,$$

$$\alpha_t | \alpha_{t-1}, \cdots, \alpha_{t-p} \sim \text{Normal } (\phi_1 \alpha_{t-1} + \cdots + \phi_p \alpha_{t-p}, \sigma^2),$$

$$\tilde{\alpha}_t | \tilde{\alpha}_{t-1}, \cdots, \tilde{\alpha}_{t-q} \sim \text{Normal } (\tilde{\phi}_1 \tilde{\alpha}_{t-1} + \cdots + \tilde{\phi}_q \tilde{\alpha}_{t-q}, \tilde{\sigma}^2),$$

subject to the constraint $0 \leq w_{3t} \leq min(y_t, z_t), t = 1, 2, \cdots, n$. The likelihood function of this model is obtained by

$$L(\theta, y, z) = \int \sum_{w_{3t}=0}^{min(y_t, z_t)} \prod_{t=1}^{n} f(y_t, z_t | w_{3t}, \alpha_t) h(w_{3t}) G(\alpha, \tilde{\alpha}) d\alpha d\tilde{\alpha},$$

where

$$G(\alpha, \tilde{\alpha}) = g(\alpha_t | \alpha_{t-1}, \cdots, \alpha_{t-p})\, \tilde{g}(\tilde{\alpha}_t | \tilde{\alpha}_{t-1}, \cdots, \tilde{\alpha}_{t-q}) g_0(\alpha_0, \cdots, \alpha_{1-p}) \tilde{g}_0(\tilde{\alpha}_0, \cdots, \tilde{\alpha}_{1-q}),$$

$\theta = (\theta_1, \theta_2, \theta_3), \theta_1 = (\beta_1, \beta_2, \beta_3)$ denotes the parameters of the fixed effects, the parameters of the latent process $\{\alpha_t\}$ are denoted by $\theta_2 = (\phi_1, \cdots, \phi_p, \sigma)$, and finally $\theta_3 = (\tilde{\phi}_1, \cdots, \tilde{\phi}_q, \tilde{\sigma})$ denotes the parameters of the latent process $\{\tilde{\alpha}_t\}$.$\alpha_0, \cdots, \alpha_{1-p}$ and $\tilde{\alpha}_0, \cdots, \tilde{\alpha}_{1-q}$ are the initial conditions of AR(p) and AR(q) processes, respectively.

The posterior distribution of θ conditional on the data $(y, z)^{(K)}$ is given by

$$\pi_K(\theta|y, z) = \frac{[L(\theta, y, z)]^K \pi(\theta)}{C(K, y, z)},$$

where $\pi(\theta)$ is the prior distribution of the parameters, and

$$C(K, y, z) = \int [L(\theta, y, z)]^K \pi(\theta) d\theta,$$

is the normalizing constant.

5.3 Numerical studies

A simulation study was conducted to check the performance of the DC method when the data is simulated from BP1 and BP2 models. We used 500 Monte Carlo realizations from each model with sample size of 500 in each realization. *dclone, rjags* and *coda* packages from R and *Jags* software were used to do these simulations.

5.3.1 Experiment 1: BP1 model

In this experiment we considered two cases:

1. Case 1: BP1 with AR(1) latent process.

2. Case 2: BP1 with AR(2) latent process.

The true values for the parameters $\lambda_{1t}, \lambda_{2t}$ and λ_{3t} in Case 1 are as follows:

$$ln\lambda_{1t} = 2 + x_t,\ ln\lambda_{2t} = 1.5 - 0.5x_t \text{ and } ln\lambda_{3t} = -1 + 2x_t + \alpha_t,$$

the latent process $\{\alpha_t\}$ is given by $\alpha_t = -0.5\alpha_{t-1} + \epsilon_t$, where $\epsilon_t \sim Normal(0, 0.7^2)$ and the explanatory variable x_t was drawn from standard normal distribution, whereas the true values for the parameters $\lambda_{1t}, \lambda_{2t}$ and λ_{3t} in Case 2 are given as follows:

$$ln\lambda_{1t} = 1.5 + 0.8x_t,\ ln\lambda_{2t} = 1 - 0.5\ sin\frac{\pi t}{3} - cos\frac{\pi t}{3} \text{ and } ln\lambda_{3t} = 1 + \alpha_t,$$

the latent process $\{\alpha_t\}$ is given by $\alpha_t = \alpha_{t-1} - 0.7\alpha_{t-2} + \epsilon_t$, where $\epsilon_t \sim Normal(0, 0.5^2)$ and x_t is standard normal random variable.

The following priors were used in our simulations: normal distribution with mean 0 and variance 10^3 for fixed effects parameters, log normal distribution with mean 0 and variance 1 for the inverse of the variance component. For the latent process AR(1), we used uniform prior distribution on the interval $(-0.99, 0.99)$ for ϕ_1, and for ϕ_1, ϕ_2 in the latent process AR(2) we used normal prior distribution with mean 0 and variance 10^4. Furthermore, we assigned normal priors distributions with mean 0 and variance 1 for the initial condition parameters in both processes.

In each case we set the following: burn-in period of 2000, three parallel MCMC chains and 5000 values to generate from the posterior distribution from each chain.

Tables 5.1 and 5.3 show the real values of the parameters, the empirical means, the empirical standard deviations, MSE and DC standard errors with ($K = 3$). Both Tables show that the true value of the parameter is very close to the estimated value, and the DC standard errors and the empirical standard deviations are in very good agreement.

Tables 5.2 and 5.4 show the real values of the parameters, the empirical means, the empirical standard deviations, MSE and Bayes MCMC standard errors with

($K = 1$). Almost we have similar results for both DC method and Bayes MCMC method except for estimating ϕ_1 in AR(1) process and ϕ_1, ϕ_2 in AR(2) process, DC method gives better unbiased estimate for these parameters than Bayes MCMC method.

Table 5.5 shows the percentage of coverage of 95% confidence interval in BP1 model with AR(1) latent process and BP1 model with AR(2) latent process.

Table 5.1: Estimation of BP1 model parameters with AR(1) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_{10}	2.00	1.9989	0.0005	0.0229	0.0228
β_{11}	1.00	1.0013	0.0002	0.0146	0.0146
β_{20}	1.50	1.4982	0.0018	0.0429	0.0402
β_{21}	-0.50	-0.5004	0.0010	0.0309	0.0310
β_{30}	-1.00	-1.0285	0.0776	0.2773	0.2683
β_{31}	2.00	2.0140	0.0301	0.1730	0.1661
σ	0.70	0.6948	0.0080	0.0893	0.0889
ϕ_1	-0.50	-0.4612	0.0294	0.1671	0.1487

Table 5.2: Estimation of BP1 model parameters with AR(1) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_{10}	2.00	1.9998	0.0005	0.0229	0.0228
β_{11}	1.00	1.0011	0.0002	0.0146	0.0146
β_{20}	1.50	1.5008	0.0019	0.0431	0.0400
β_{21}	-0.50	-0.4981	0.0010	0.0312	0.0310
β_{30}	-1.00	-1.0756	0.0880	0.2871	0.2748
β_{31}	2.00	2.0391	0.0331	0.1778	0.1702
σ	0.70	0.7157	0.0079	0.0876	0.0898
ϕ_1	-0.50	-0.4214	0.0343	0.1679	0.1559

Table 5.3: Estimation of BP1 model parameters with AR(2) latent process using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_{10}	1.50	1.5004	0.0017	0.0417	0.0426
β_{11}	0.80	0.8005	0.0006	0.0251	0.0259
β_{20}	1.00	0.9925	0.0058	0.0760	0.0719
β_{21}	-0.50	-0.5028	0.0031	0.0557	0.0536
β_{22}	-1.00	–1.0079	0.0054	0.0732	0.0681
β_{30}	1.00	0.9943	0.0068	0.0825	0.0806
σ	0.50	0.5032	0.0024	0.0493	0.0476
ϕ_1	1.00	0.9883	0.0039	0.0612	0.0591
ϕ_2	-0.70	-0.6913	0.0034	0.0576	0.0524

Table 5.4: Estimation of BP1 model parameters with AR(2) latent process using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_{10}	1.50	1.4998	0.0017	0.0416	0.0427
β_{11}	0.80	0.8007	0.0006	0.0250	0.0259
β_{20}	1.00	0.9896	0.0059	0.0758	0.0723
β_{21}	-0.50	-0.5039	0.0031	0.0556	0.0534
β_{22}	-1.00	-1.0099	0.0054	0.0733	0.0684
β_{30}	1.00	0.9915	0.0068	0.0820	0.0816
σ	0.50	0.5113	0.0025	0.0488	0.0484
ϕ_1	1.00	0.9808	0.0042	0.0615	0.0601
ϕ_2	-0.70	-0.6836	0.0037	0.0582	0.0534

Table 5.5: Percentage of coverage of a 95% confidence interval in BP1 model

Parameters	AR(1) latent process	Parameters	AR(2) latent process
β_{10}	0.948	β_{10}	0.956
β_{11}	0.948	β_{11}	0.952
β_{20}	0.928	β_{20}	0.934
β_{21}	0.948	β_{21}	0.936
β_{30}	0.944	β_{22}	0.918
β_{31}	0.936	β_{30}	0.948
σ	0.946	σ	0.948
ϕ_1	0.918	ϕ_1	0.938
-	-	ϕ_2	0.932

5.3.2 Experiment 2: BP2 model

In this experiment we also considered two cases:

1. Case 1: BP2 with two AR(1) latent processes.

2. Case 2: BP2 with two AR(2) latent processes.

The true values for the parameters $\lambda_{1t}, \lambda_{2t}$ and λ_{3t} in Case 1 are as follows:

$$ln\lambda_{1t} = 1 + 0.9x_t + \alpha_t, \quad ln\lambda_{2t} = 0.8 + 0.5x_t + \tilde{\alpha}_t$$

and

$$ln\lambda_{3t} = -1.4 + 2z_t,$$

the two latent processes are given by

$$\alpha_t = -0.6\alpha_{t-1} + \epsilon_t, \text{ and } \tilde{\alpha}_t = 0.6\tilde{\alpha}_{t-1} + \tilde{\epsilon}_t,$$

where $\epsilon_t \sim Normal(0, 0.8^2)$, and $\tilde{\epsilon}_t \sim Normal(0, 0.4^2)$, the explanatory variable x_t is a standard normal random variable and z_t is uniform random variable on the interval $(0, 2)$.

The true values for the parameters $\lambda_{1t}, \lambda_{2t}$ and λ_{3t} in Case 2 are as follows:

$$ln\lambda_{1t} = 1 + 0.7x_t + \alpha_t, \quad ln\lambda_{2t} = 1.5 - 0.8x_t + \tilde{\alpha}_t$$

and

$$ln\lambda_{3t} = 0.5 + x_t,$$

the latent processes are

$$\alpha_t = -0.4\alpha_{t-1} - 0.5\alpha_{t-2} + \epsilon_t, \quad \text{where } \epsilon_t \sim Normal(0, 0.6^2),$$

and

$$\tilde{\alpha}_t = \tilde{\alpha}_{t-1} - 0.7\tilde{\alpha}_{t-2} + \tilde{\epsilon}_t, \quad \text{where } \tilde{\epsilon}_t \sim Normal(0, 0.5^2)$$

and x_t is standard normal random variable.

The following priors were used in our simulations: $Normal(0, 10^3)$ for β_1, β_2 and β_3, *log* $Normal(0, 1)$ for σ and $\tilde{\sigma}$, $uniform(-0.99, 0.99)$ for ϕ_1 and $\tilde{\phi}_1$ in AR(1) processes, $Normal(0, 10^4)$ for ϕ_1, ϕ_2 and $\tilde{\phi}_1, \tilde{\phi}_2$ in AR(2) processes, and finally, $Normal(0, 1)$ for the initial condition parameters in both processes.

In each case we set the following: burn-in period of 2000, three parallel MCMC chains with 5000 iterations from each chain.

Tables 5.6 and 5.8 report the real values of the parameters, the empirical means, the empirical standard deviations, MSE and DC standard errors with $(K = 3)$ and $(K = 5)$, respectively. Both Tables show that the true value of the parameter is very close to the estimated value, and the DC standard errors and the empirical standard deviations are in very good agreement.

Tables 5.7 and 5.9 show the real values of the parameters, the empirical means, the empirical standard deviations, MSE and Bayes MCMC standard errors with $(K = 1)$. Almost we have similar results for both DC method and Bayes MCMC method.

Table 5.10 shows the percentage of coverage of 95% confidence interval in BP2 model with AR(1) latent process and BP2 model with AR(2) latent process.

Table 5.6: Estimation of BP2 model parameters with two AR(1) latent processes using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_{10}	1.00	0.9982	0.0046	0.0677	0.0664
β_{11}	0.90	0.9000	0.0037	0.0604	0.0555
β_{20}	0.80	0.7975	0.0061	0.0780	0.0764
β_{21}	0.50	0.4975	0.0018	0.0429	0.0427
β_{30}	-1.40	-1.4070	0.0370	0.1923	0.1856
β_{31}	2.00	2.0028	0.0114	0.1068	0.1036
σ	0.80	0.7991	0.0028	0.0528	0.0509
ϕ_1	-0.60	-0.5947	0.0028	0.0523	0.0505
$\widetilde{\sigma}$	0.40	0.3996	0.0032	0.0565	0.0565
$\widetilde{\phi}_1$	0.60	0.5681	0.0122	0.1033	0.1035

Table 5.7: Estimation of BP2 model parameters with two AR(1) latent processes using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_{10}	1.00	0.9927	0.0046	0.0678	0.0671
β_{11}	0.90	0.9019	0.0037	0.0605	0.0559
β_{20}	0.80	0.7909	0.0063	0.0785	0.0778
β_{21}	0.50	0.4982	0.0019	0.0430	0.0430
β_{30}	-1.40	-1.4013	0.0365	0.1911	0.1849
β_{31}	2.00	1.9997	0.0112	0.1060	0.1031
σ	0.80	0.8059	0.0028	0.0528	0.0516
ϕ_1	-0.60	-0.5910	0.0029	0.0527	0.0510
$\widetilde{\sigma}$	0.40	0.4075	0.0031	0.0553	0.0558
$\widetilde{\phi_1}$	0.60	0.5545	0.0133	0.1058	0.1045

Table 5.8: Estimation of BP2 model parameters with two AR(2) latent processes using DC method

Parameter	Real value	MLE estimator	MSE	Empirical SD	DC SE
β_{10}	1.00	0.9936	0.0056	0.0749	0.0714
β_{11}	0.70	0.7033	0.0022	0.0472	0.0479
β_{20}	1.50	1.4945	0.0047	0.0685	0.0647
β_{21}	-0.80	-0.8031	0.0023	0.0482	0.0466
β_{30}	0.50	0.4985	0.0112	0.1057	0.1004
β_{31}	1.00	1.0012	0.0030	0.0545	0.0534
σ	0.60	0.5950	0.0027	0.0522	0.0496
ϕ_1	-0.40	-0.4026	0.0074	0.0861	0.0802
ϕ_2	-0.50	-0.5061	0.0045	0.0668	0.0697
$\widetilde{\sigma}$	0.50	0.4966	0.0012	0.0346	0.0356
$\widetilde{\phi}_1$	1.00	1.0006	0.0026	0.0509	0.0497
$\widetilde{\phi}_2$	-0.70	-0.7014	0.0020	0.0444	0.0446

Table 5.9: Estimation of BP2 model parameters with two AR(2) latent processes using Bayes MCMC.

Parameter	Real value	Bayes estimator	MSE	Empirical SD	Bayes SE
β_{10}	1.00	0.9857	0.0060	0.0759	0.0722
β_{11}	0.70	0.7073	0.0023	0.0478	0.0487
β_{20}	1.50	1.4895	0.0049	0.0690	0.0652
β_{21}	-0.80	-0.8052	0.0024	0.0487	0.0470
β_{30}	0.50	0.5012	0.0111	0.1057	0.1004
β_{31}	1.00	0.9989	0.0029	0.0543	0.0533
σ	0.60	0.6030	0.0027	0.0523	0.0501
ϕ_1	-0.40	-0.3953	0.0073	0.0853	0.0808
ϕ_2	-0.50	-0.4978	0.0044	0.0660	0.0704
$\widetilde{\sigma}$	0.50	0.5052	0.0012	0.0347	0.0363
$\widetilde{\phi}_1$	1.00	0.9940	0.0027	0.0513	0.0507
$\widetilde{\phi}_2$	-0.70	-0.6945	0.0021	0.0453	0.0456

Table 5.10: Percentage of coverage of a 95% confidence interval in BP2 model

Parameters	AR(1) latent processes	Parameters	AR(2) latent processes
β_{10}	0.942	β_{10}	0.932
β_{11}	0.920	β_{11}	0.954
β_{20}	0.940	β_{20}	0.932
β_{21}	0.942	β_{22}	0.936
β_{30}	0.942	β_{30}	0.936
β_{31}	0.938	β_{31}	0.952
σ	0.944	σ	0.948
ϕ_1	0.946	ϕ_1	0.922
$\widetilde{\sigma}$	0.946	ϕ_2	0.944
$\widetilde{\phi_1}$	0.942	$\widetilde{\sigma}$	0.946
-	-	$\widetilde{\phi_1}$	0.942
-	-	$\widetilde{\phi_2}$	0.940

5.4 Real data application

5.4.1 Asthma visits by asthma type

In this section, we will analyze daily counts of emergency department visits due to asthma in the Canadian province of Ontario during the period January 1st, 2010 till December 29th, 2016 (sample size = 2555). The data set was obtained from the Canadian Institute for Health Information and it consisted of daily counts of visits

for two types of asthma (codes: J4500 (Predominantly allergic asthma without stated status asthmaticus) and J4590 (Asthma, unspecified, without stated status asthmatics)). The data are summarised in Figures 5.1, 5.2 and Table 5.11. It is clear from these figures that there is some form of seasonal pattern with higher activity occurring in September and October for type J4500 and in September, October and December for type J4590. Also, there is suggestion to include effects of weekends (Saturday and Sunday) because during weekends general practitioners (private physicians) are less available and people tend to rely more on emergency departments of hospitals when asthma attack occurs.

Table 5.11: A statistical summary of asthma dataset for ICD codes J4590 and J4500

	Asthma J4500	Asthma J4590
Mean	56.9	96.0
Variance	551.9	536.3
Standard deviation	23.5	23.2
Minimum	12	47
Maximum	268	221

Figure 5.1: Asthma J4500 presentations

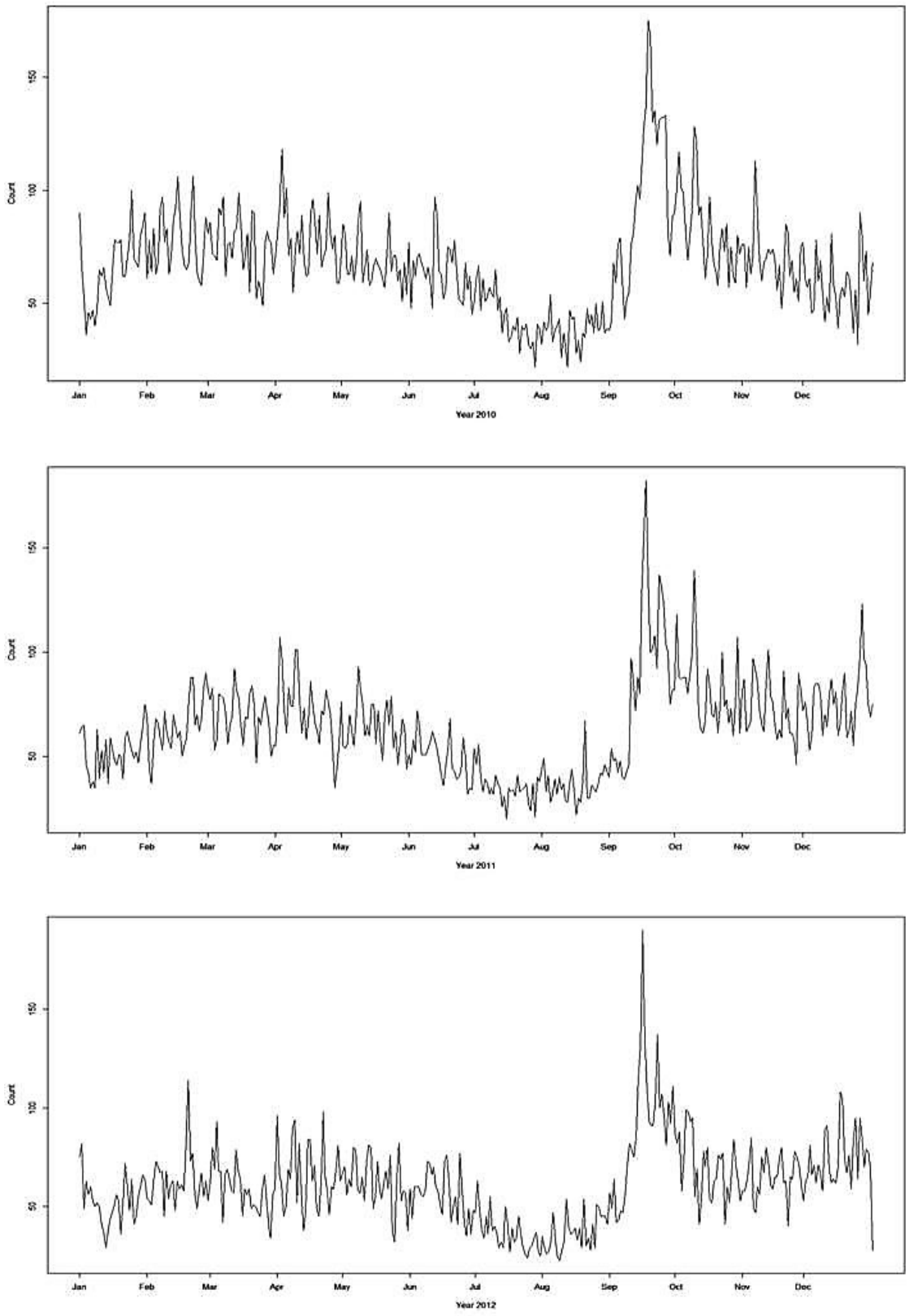
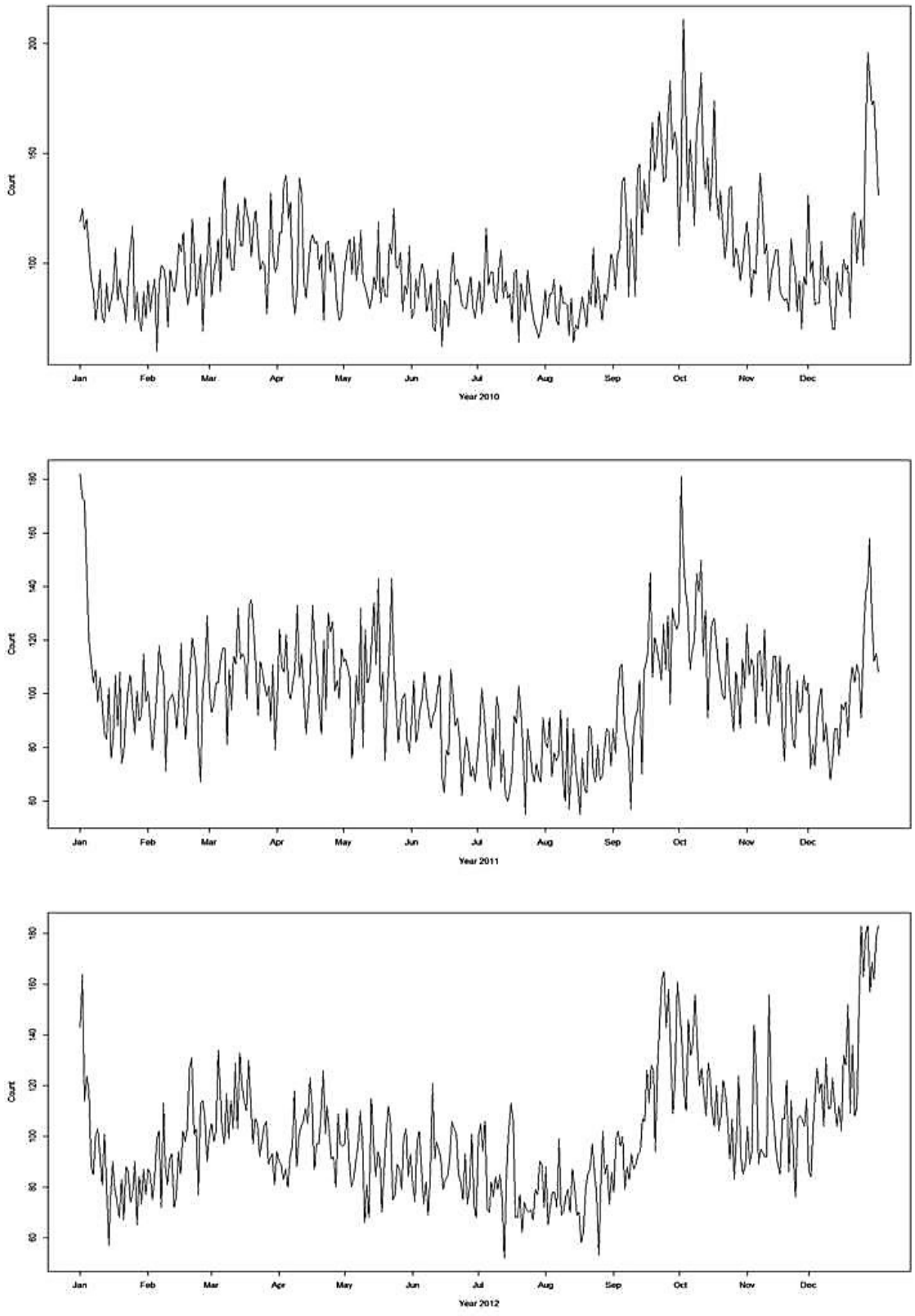


Figure 5.2: Asthma J4590 presentations



For this data set, We fitted the following model:

$$ln\lambda_{it} = \beta_{i0} + \sum_{j=1}^{6} \beta_{ij} x_{jt}, \quad (i = 1, 2)$$

and

$$ln\lambda_{3t} = \beta_{30} + \alpha_t, \quad (t = 1, \cdots, 2555),$$

where $\alpha_t = \phi_1 \alpha_{t-1} + \epsilon_t$ and $\epsilon_t \sim Normal(0, \sigma^2)$. The explanatory variables are as follows: $x_{1t} = \frac{t}{2555}$ to include trend; x_{2t} to include weekend effect; $x_{3t} = cos\frac{2\pi t}{365}$; $x_{4t} = sin\frac{2\pi t}{365}$; $x_{5t} = cos\frac{4\pi t}{365}$ and $x_{6t} = sin\frac{4\pi t}{365}$ to include seasonal effects. The dependent variables are, of course, counts of daily visits by people with asthma types J4500 and J4590 in the province of Ontario in the period mentioned above.

Table 5.12 reports the model parameter estimates and the corresponding standard errors. For this application, the number of colons was k=5, burn-in period of 20000 iterations, three MCMC chains with 10000 iterations each. Furthermore, the following priors were used: $Normal(0, 10^3)$ for the fixed random effects; *log* $Normal(0, 1)$ for the variance components and *uniform*$(-0.99, 0.99)$ for the correlation component.

To check the convergence of the DC approach, we calculated the largest eigenvalue of the posterior variance-covariance matrix, the mean square error and the correlation-like fit statistic. The values are 0.0026, 0.0176 and 0.0000, respectively. The values reflect the degenerateness of the posterior distribution and that the normal approximation is adequate. Also, Brooks-Gelman statistic was $\widehat{R} = 1.0905$, indicating MCMC chain convergence.

In order to examine the convergence of the DC approach, we provided plots of the posterior densities of the model parameters (Figure 5.3) and their trace plots (Figure 5.4). The posterior densities look appropriately normal, and the trace

plots suggest that the samples of all parameters got mixed well, but the samples of the cross correlation component show less quality of mixing.

Table 5.12: Estimates and their standard errors from analysis of asthma data for types J4500 and J4590

Parameter	MLE	DC SE	z-value	Parameter	MLE	DC SE	z-value
β_{10}	3.9704	0.0123	322.80**	β_{20}	4.4774	0.0077	581.04**
β_{11}	-0.4593	0.0219	-20.97**	β_{21}	-0.1305	0.0127	10.28**
β_{12}	0.1258	0.0079	15.92**	β_{22}	0.0432	0.0053	8.15**
β_{13}	0.1659	0.0101	16.43**	β_{23}	0.1199	0.0055	21.80**
β_{14}	0.0157	0.0090	1.74	β_{24}	0.0328	0.0055	5.96**
β_{15}	-0.1733	0.0092	-18.84**	β_{25}	-0.0570	0.0053	10.75**
β_{16}	-0.1297	0.0092	-14.10**	β_{26}	-0.0519	0.0054	9.61**
ϕ_1	0.9286	0.0103	90.16**	β_{30}	1.8229	0.1142	15.96**
σ	0.4804	0.0210	22.88**	-	-	-	

** indicates significant at 0.001 level

Figure 5.3: The posterior densities of BP1 model parameters

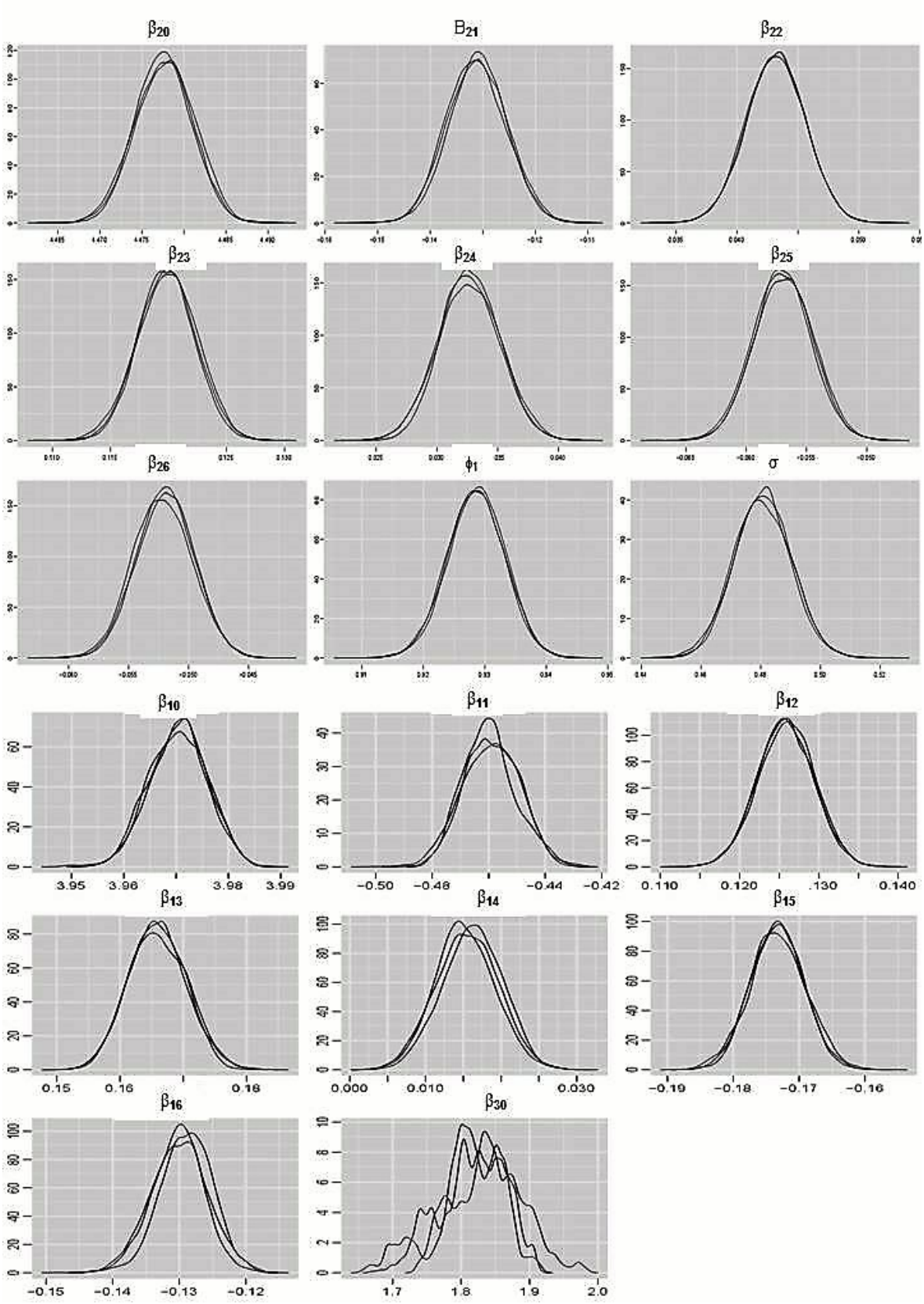


Figure 5.4: The trace plots of BP1 model parameters

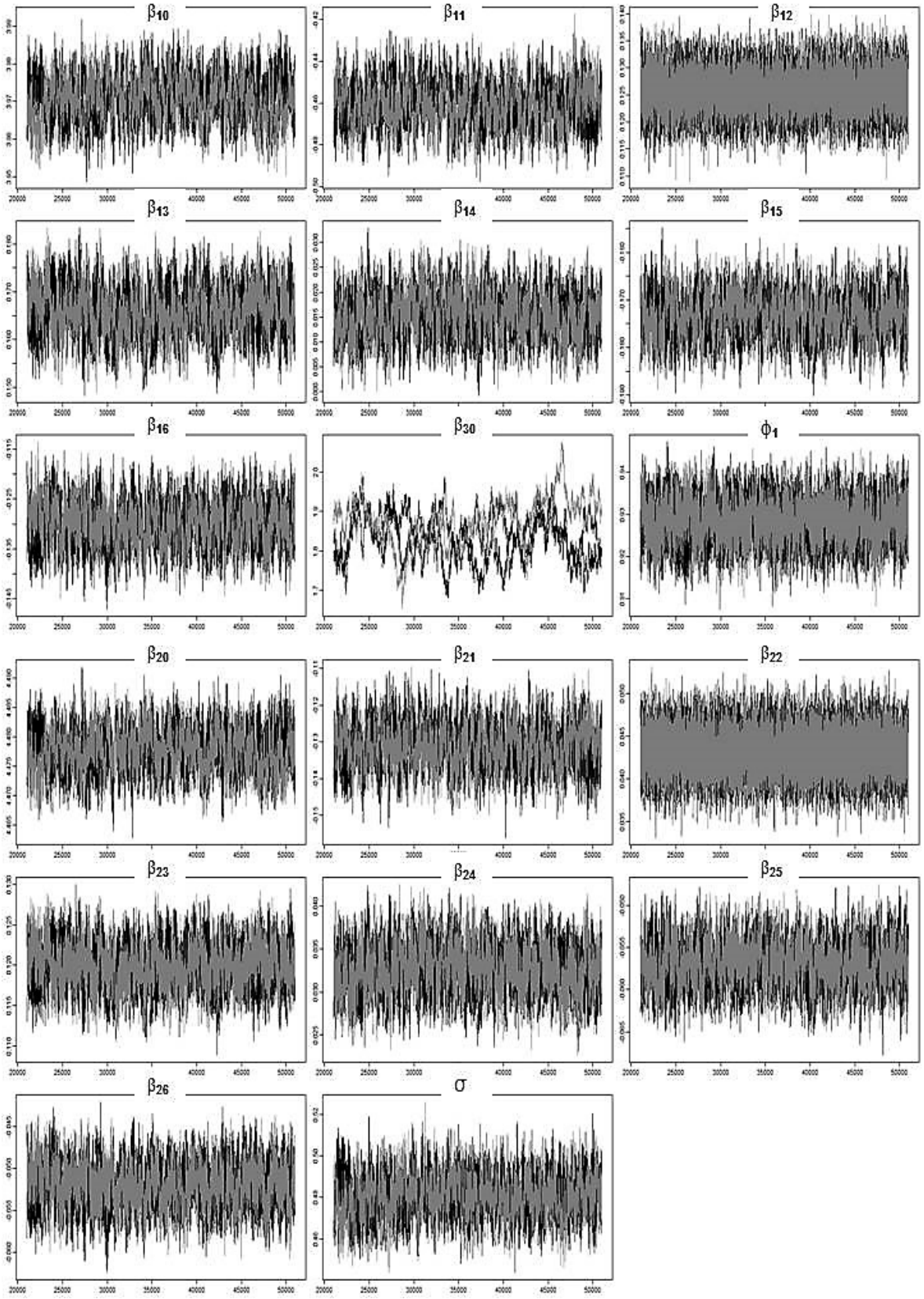


Table 5.12 reveals that there is a significant trend effect indicating decrease in the counts of both asthma types. Also, weekend effect is highly significant variable for both asthma types. Fourier series terms to model seasonal pattern in the data are statistically significant except $sin\frac{2\pi t}{365}$ for asthma type J4500. Furthermore, the cross correlation component is significant and the parameters of the latent process are also significant. It is noteworthy that the strength of the various regression coefficients are not same for the two types of asthma. For instance, the effect of weekend as compared to week days is 0.1258 for J4500 while such effect is 0.0432 for J4590. This is an indication that the two asthma types do not behave the same way and hence, a bivariate modeling, or in general, regressions models that take into account the type variable, are necessary.

5.4.2 Asthma visits by age group

Another scenario where a bivariate count data may arise is when one considers emergency department visits due to asthma for different age groups. Here, for the sake of illustration, we will analyze daily counts of emergency department visits due to asthma in the Canadian province of Ontario during the period January 1st, 2010 till December 10th, 2015 (sample size = 2170) for children in the age groups (0-9) and (10-19) years. The data set was obtained from the Canadian Institute for Health Information.

Figures 5.5 and 5.6 show time series plots of these data, while Table 5.13 provides summary statistics. It is clear from these figures that there is some form of seasonal pattern with higher activity occurring in the fall (September-November) for both age groups. Also, there is suggestion to include model terms for weekend effect (Saturday and Sunday) for the same reasons as stated in the previous section.

Table 5.13: A statistical summary of asthma dataset

	Age group (0-9) years	Age group (10-19) years
Mean	48.9	22.3
Variance	453.0	87.8
Standard deviation	21.3	9.4
Minimum	11	5
Maximum	237	91

Figure 5.5: Asthma presentations for age group (0-9) years

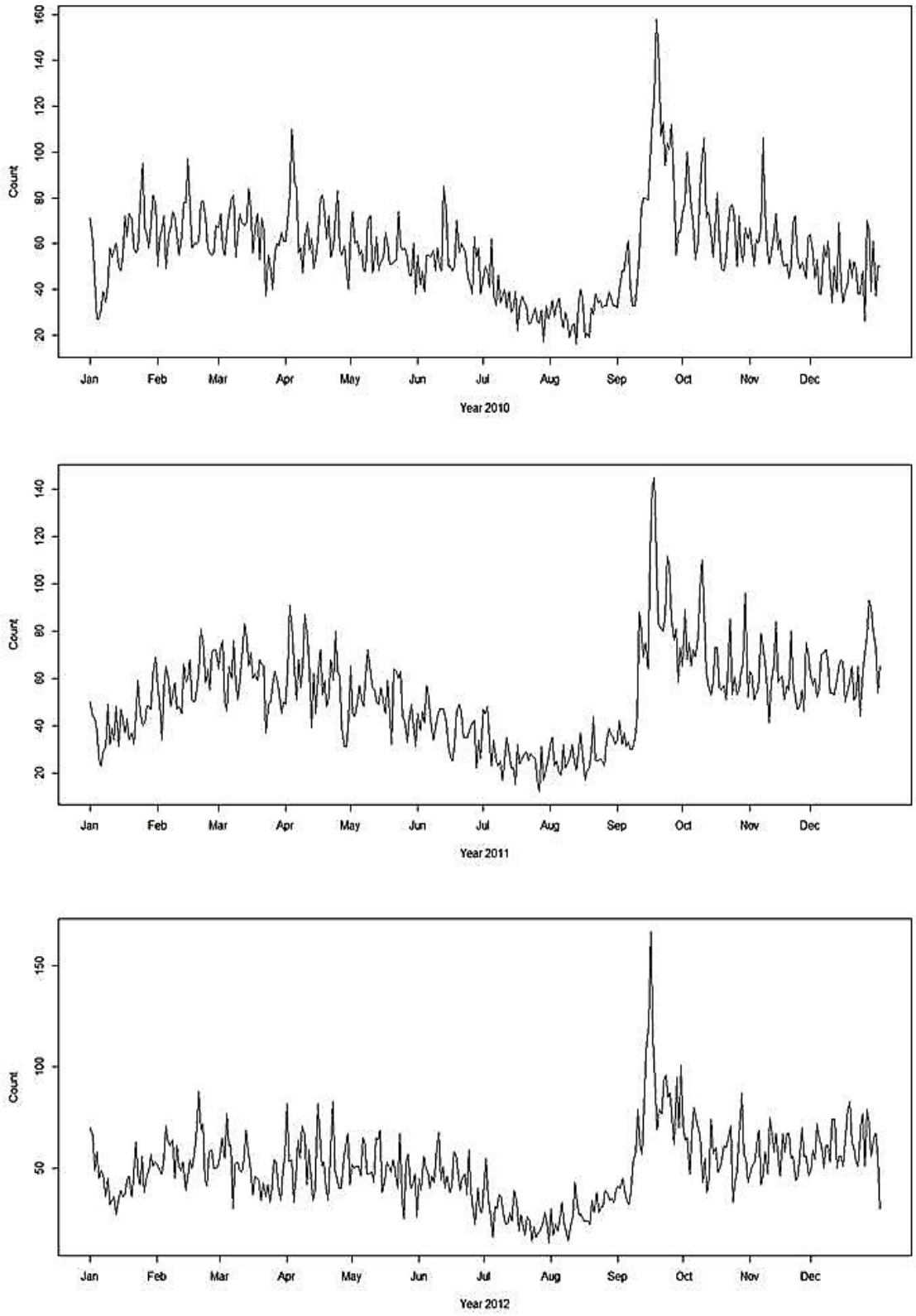
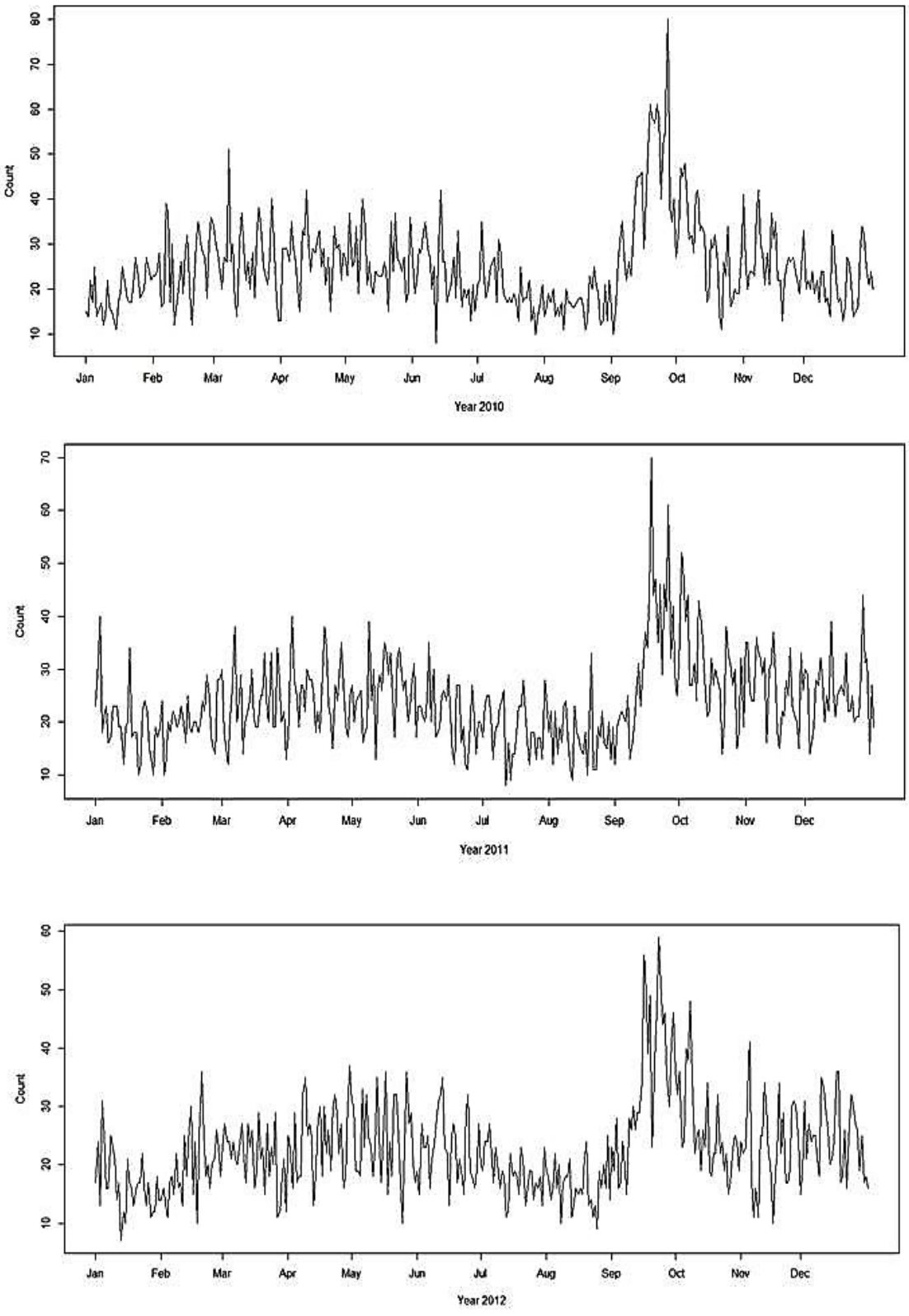


Figure 5.6: Asthma presentations for age group (10-19) years



The following model is found useful in our exploration of the data:

$$ln\lambda_{1t} = \beta_{10} + \sum_{j=1}^{6} \beta_{1j} x_{jt} + \alpha_t, \quad ln\lambda_{2t} = \beta_{20} + \sum_{j=1}^{6} \beta_{2j} x_{jt} + \widetilde{\alpha}_t$$

and

$$ln\lambda_{3t} = \beta_{30},$$

where $\alpha_t \sim Normal(0, \sigma^2)$, $\widetilde{\alpha}_t = \widetilde{\phi}_1 \widetilde{\alpha}_{t-1} + \epsilon_t$ and $\epsilon_t \sim Normal(0, \widetilde{\sigma}^2)$. The explanatory variables are as follows: $x_{1t} = \frac{t}{2170}$ to include trend; x_{2t} to include weekend effect; $x_{3t} = cos\frac{2\pi t}{365}$; $x_{4t} = sin\frac{2\pi t}{365}$; $x_{5t} = cos\frac{4\pi t}{365}$ and $x_{6t} = sin\frac{4\pi t}{365}, (t = 1, \cdots, 2170)$. The dependent variables are:

1. Emergency department daily visits by children aged (0-9) years.

2. Emergency department daily visits by people aged (10-19) years.

Table 5.14 reports the model parameter estimates and the corresponding standard errors. For this application, the number of clones was k=5, burn-in period of 30000 iterations, two MCMC chains with 10000 iterations each. Furthermore, the following priors were used: $Normal(0, 10^3)$ for fixed random effects; *log Normal*$(0, 1)$ for variance components and *uniform*$(-0.99, 0.99)$ for correlation component.

To check the convergence of the DC approach, we calculated the largest eigenvalue of the posterior variance-covariance matrix, the mean square error and the correlation-like fit statistic. The values are 0.0019, 0.0091 and 0.0002, respectively. The values reflect the degenerateness of the posterior distribution and that the normal approximation is adequate. Also, Brooks-Gelman statistic, $\widehat{R} = 1.0569$, indicating MCMC chain convergence.

For further investigate the behaviour of the convergence of the DC approach,

Figure 5.7 provides a plot of the posterior densities of the model parameters which look appropriately normal, and Figure 5.8 presents trace plots for model parameters and it suggests that the samples of the parameters got mixed well except the samples of β_{01}, β_{02} and β_{03} show less quality of mixing.

Table 5.14: Estimates and their standard errors from analysis of asthma data for age groups (0-9) and (10-19) years

Parameter	MLE	DC SE	z-value	Parameter	MLE	DC SE	z-value
β_{10}	3.7369	0.0240	155.70**	β_{20}	2.6323	0.0650	40.50**
β_{11}	-0.4809	0.0303	-15.87**	β_{21}	-0.5692	0.0907	-6.28**
β_{12}	0.1740	0.0197	8.83**	β_{22}	0.0679	0.0225	3.02*
β_{13}	0.2520	0.0132	19.09**	β_{23}	-0.0090	0.0378	-0.24
β_{14}	-0.0116	0.0118	-0.98	β_{24}	-0.1657	0.0369	-4.49**
β_{15}	-0.3152	0.0133	-23.70**	β_{25}	-0.3343	0.0373	-8.96**
β_{16}	-0.1963	0.0128	-15.34**	β_{26}	-0.2817	0.0377	-7.47**
σ	0.3473	0.0086	40.38**	$\widetilde{\sigma}$	0.2482	0.0167	14.86**
β_{30}	2.2915	0.0546	41.97**	$\widetilde{\phi}_1$	0.7806	0.0251	31.10**

* indicates significant at 0.01 level, ** indicates significant at 0.001 level

Figure 5.7: The posterior densities of BP2 model parameters

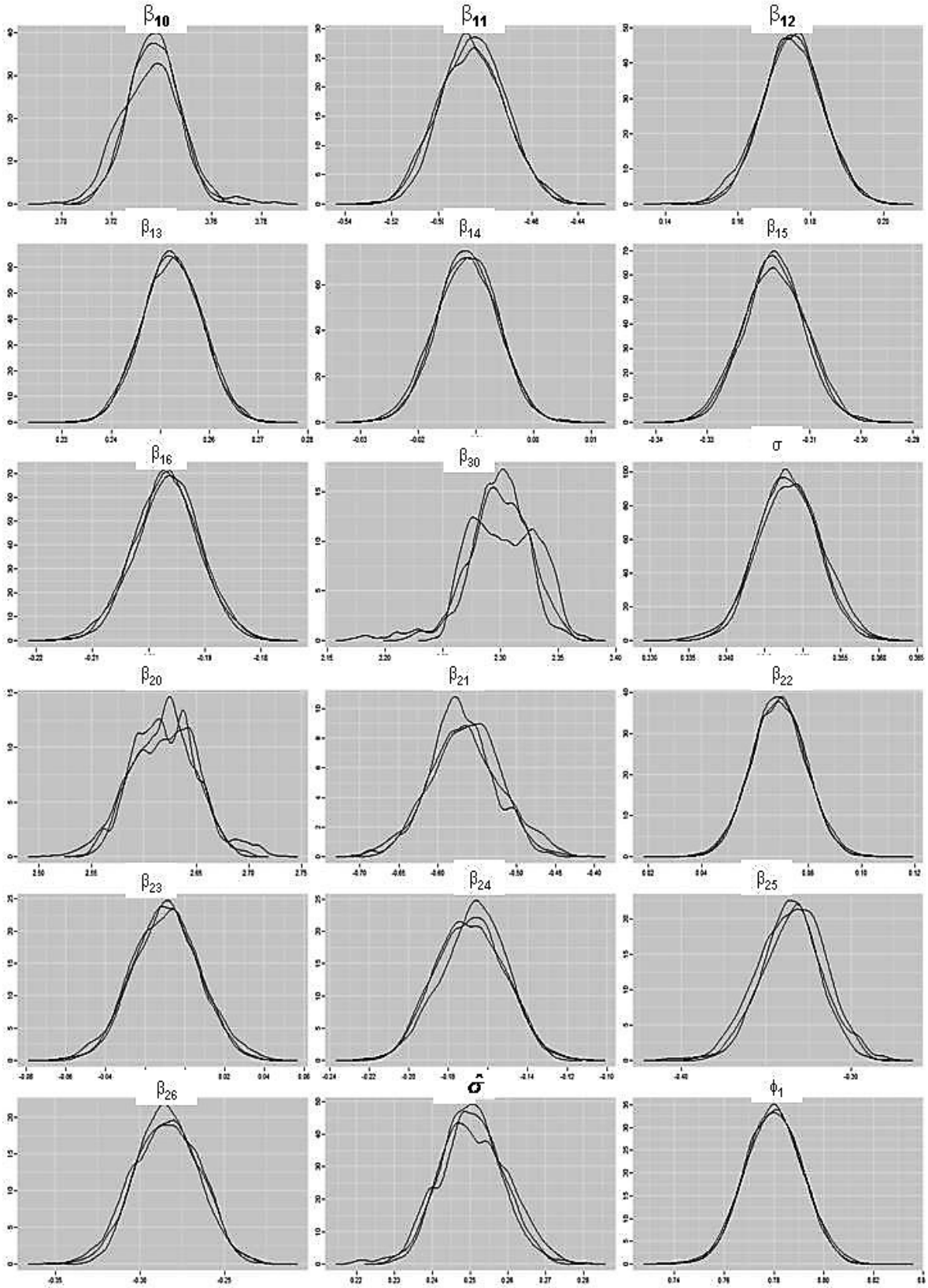


Figure 5.8: The trace plots of BP2 model parameters

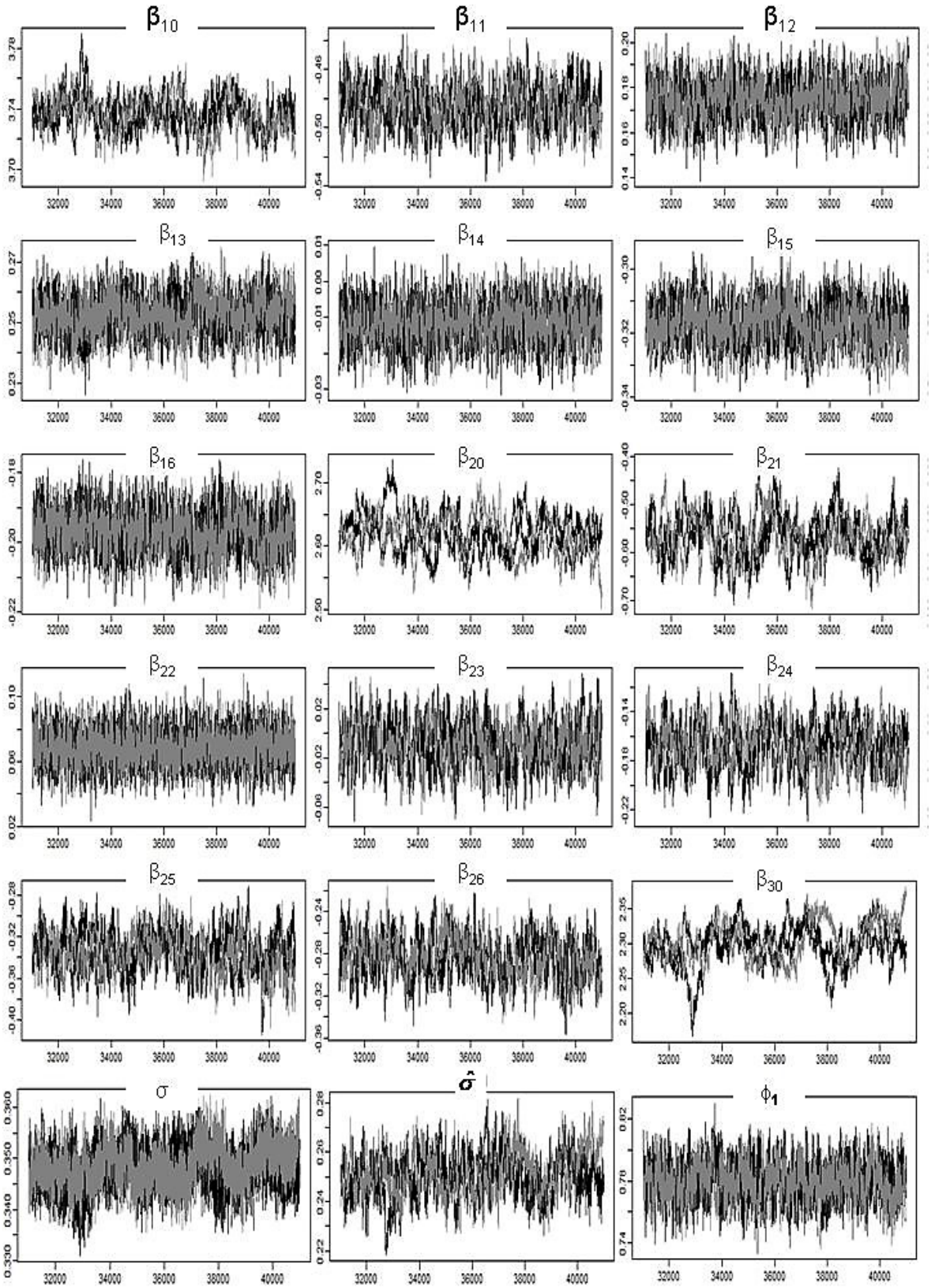


Table 5.14 reveals that weekend effect is highly significant variable for both age groups but with larger effect on the group (0-9) years. Also, there is significant trend effect indicating decrease in the counts of both age groups. In addition, Fourier series terms to model seasonal pattern in the data are statistically significant except $sin\frac{2\pi t}{365}$ for age group (0-9) years and $cos\frac{2\pi t}{365}$ for age group (10-19) years.

The variance components for both age groups are statistically significant. This means that the daily number of asthma visitors to the emergency department of hospitals for both age groups express significant heterogeneity. Also, there is significant correlation between observations for the age group (10-19) years, and the parameter of cross correlation β_{30} is also statistically significant.

Chapter 6

Summary and future research

6.1 Summary and Outline of Contributions

Parameter-driven models for time series of counts are attractive because the regression coefficients therein are interpretable in the same way that a generalized linear model with random effects is interpreted. A major difficulty posed by these models, however, comes from their computational intractability. There have been many works in the literature using various computational approaches, including Bayesian, EM and Particle filtering, to estimate the parameters of these models and carry out inferences. The computational difficulty is further complicated when the counts are bivariate, or more generally multivariate, or the data has extra zeros than expected by the commonly used models.

In this dissertation, we studied three kinds of parameter-driven count models. The main goal of these models is to accommodate correlation between observations in time series data.

Namely, in Chapter 2 and 3, we presented ZIP and ZINB parameter-driven models and derived their marginal moments. Using data cloning method, we computed the maximum likelihood estimates of the models parameters and their

asymptotic variance. As an application of these models, we used two datasets: the first one is daily counts of emergency department visits for asthma type J4591 in Ontario, and the second one is the daily counts of emergency department visits for asthma of the age group (70-79) years in Ontario.

In Chapter 4, we proposed hurdle parameter-driven models for both Poisson and negative binomial and used data cloning method to find the MLE for their parameters. We applied these models on asthma datasets for the age group (70-79) years. Using AIC difference procedure we showed that NBARH parameter-driven model provides better description of the data than does PARH in both datasets.

Finally, in Chapter 5, we showed two kinds of parameter-driven models of bivariate Poisson distribution. One with one latent process and the other with two latent processes. We derived the marginal moments of the observed bivariate process, and numerical simulations were conducted on the estimation of the model's parameters. We applied the BP1 model to analyze linear and seasonal trends in the counts of daily emergency department visits due asthma with ICD codes J4500 and J4590 in the province of Ontario. Similarly, we used the the BP2 model to analyze daily visits due to asthma in the age groups (0-9) years and (10-19) years.

Here is a list of contributions in this study:

1. In Chapter 2 and 3:

   (a) We computed the moments of ZIP and ZINB parameter-driven models under the assumption of autoregressive latent process of order p.

   (b) We formulated the two models in a hierarchical form appropriate for the data cloning (DC) algorithm and carried out inferences based on the DC method.

2. In Chapter 4:

(a) We proposed new hurdle parameter-driven models for both Poisson and negative binomial counts with extra zeros.

(b) We formulated the two models in a hierarchical format and applied a DC algorithm to carry out the statistical inferences.

3. In Chapter 5:

(a) The following two new parameter-driven bivariate Poisson models were introduced:

1- BP1 model by including an AR(p) process to the cross correlation parameter of a bivariate Poisson distribution.
2- BP2 model by including two latent processes, AR(p) and AR(q), in the marginal distributions of a bivariate Poisson model.

These two models are useful, in situations where the components of a bivariate count time series have same temporal autocorrelation behavior or different temporal autocorrelation behavior.

(b) We derived the moments of these new models, formulated them in hierarchical specifications and used the DC method to obtain the MLEs of their parameters.

6.2 Future study

In this dissertation we only considered autoregressive latent processes to model temporal correlation. For more complicated correlation structure in the data, it may be desirable to use a more general formula for the latent process like mixed

autoregressive/moving-average process, mixed ARMA, which contains both autoregressive terms and moving average terms.

In future, it is important to develop model diagnostic tools based on residuals. Such tools do not exist for models estimated via DC algorithm. However, in general it is not easy to define and provide asymptotic results for residuals in the area of mixed effects models, which includes the parameter-driven models as special cases (Song [61])

When estimating the parameters of ZINB and NBARH parameter-driven models, we couldn't use data cloning method to estimate the dispersion parameter r directly. We computed the parameters of the model for different values of r, then using AIC difference, we chose the estimate of r that gives us the smallest AIC. This procedure is time consuming. It will be a major improvement if we can find a procedure that can compute r directly without using AIC difference.

In Chapter 5, we presented a detailed description of two parameter-driven bivariate Poisson models. The proposed models can only account for positive correlation. From practical point of view, it is desirable to have flexible models that allows for both positive and negative correlation between observations.

Also, we focused in Chapter 5 on bivariate Poisson which can be generalized to parameter-driven multivariate Poisson model.

Appendices

A Derivation of the likelihood function

In this appendix, we will outline the derivation of the likelihood function used throughout the thesis. Consider a stationary autoregressive process of order p, AR(p), such that

$$\alpha_t = \phi_1 \alpha_{t-1} + \phi_2 \alpha_{t-2} + \cdots + \phi_p \alpha_{t-p} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a normal random process with mean zero and variance σ. Conditioning on α_t, suppose Y_t is a sequence of independent counts with conditional probability density $p(y_t|\alpha_t)$. Also, assume

$$p(y_t|\alpha_t) = p(y_t|\alpha_t, \boldsymbol{\alpha}^{(t-1)}) = p(y_t|\alpha_t, \boldsymbol{\alpha}^{(t-1)}, \mathbf{y}^{(t-1)}), \qquad t = 1, 2, \cdots$$

where $\boldsymbol{\alpha}^{(t-1)} = (\alpha_{t-1}, \cdots, \alpha_0, \alpha_{-1}, \cdots, \alpha_{t-p})$ and $\mathbf{y}^{(t-1)} = (y_{t-1}, \cdots, y_1)$. Notice that from the definition of AR(p) process we have

$$p(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p}) = p(\alpha_t|\boldsymbol{\alpha}^{(t-1)}) = p(\alpha_t|\boldsymbol{\alpha}^{(t-1)}, \mathbf{y}^{t-1)}), \qquad t = 1, 2, \cdots$$

Now, the joint likelihood of the data and the latent processes can be written

as follows:

$$p(y_1, \cdots, y_n, \alpha_n, \cdots, \alpha_1, \alpha_0, \alpha_{-1}, \cdots, \alpha_{t-p}) = p(y_n|\boldsymbol{\alpha}^{(n)}, \mathbf{y}^{(n-1)})p(\boldsymbol{\alpha}^{(n)}, \mathbf{y}^{(n-1)})$$

$$= p(y_n|\alpha_n)p(\alpha_n|\boldsymbol{\alpha}^{(n-1)}, \mathbf{y}^{(n-1)})p(\boldsymbol{\alpha}^{(n-1)}, \mathbf{y}^{(n-1)})$$

$$= p(y_n|\alpha_n)p(\alpha_n|\alpha_{n-1}, \cdots, \alpha_{n-p})p(y_{n-1}|\boldsymbol{\alpha}^{(n-1)}, \mathbf{y}^{(n-2)})p(\boldsymbol{\alpha}^{(n-1)}, \mathbf{y}^{(n-2)})$$

$$= \cdots$$

$$= \prod_{t=2}^{n} p(y_t|\alpha_t)p(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p})p(\alpha^{(1)}, y_1)$$

$$= \prod_{t=1}^{n} p(y_t|\alpha_t)p(\alpha_t|\alpha_{t-1}, \cdots, \alpha_{t-p})p(\alpha_0, \cdots, \alpha_{1-p}).$$

B Some useful propositions

The next two propositions show that the distribution of the observed processes, Y_t and Z_t, are independent of the future of the latent processes given their current values.

Proposition B.1. *Given that* $p(y_t|\alpha_t) = p(y_t|\alpha_t, \boldsymbol{\alpha}^{(t-1)})$, *where* α_t *is an AR(p) process, we have*

$$p(y_t|\alpha_t) = p(y_t|\boldsymbol{\alpha}^{(t+h)}), \qquad (t = 1, 2, \cdots) \text{ and } (h = 1, 2, \cdots)$$

Proof.

$$\begin{aligned}
p(y_t|\boldsymbol{\alpha}^{(t+h)}) =& \frac{p(\boldsymbol{\alpha}^{(t+h)}, y_t)}{p(\boldsymbol{\alpha}^{(t+h)})} \\
=& \frac{p(\alpha_{t+h}|\boldsymbol{\alpha}^{(t+h-1)}, y_t)p(\boldsymbol{\alpha}^{(t+h-1)}, y_t)}{p(\boldsymbol{\alpha}^{(t+h)})} \\
=& \frac{p(\alpha_{t+h}|\alpha_{t+h-1}, \cdots, \alpha_{t+h-p})p(\alpha_{t+h-1}|\boldsymbol{\alpha}^{(t+h-2)}, y_t)p(\boldsymbol{\alpha}^{(t+h-2)}, y_t)}{p(\boldsymbol{\alpha}^{(t+h)})} \\
=& \frac{p(\alpha_{t+h}|\alpha_{t+h-1}, \cdots, \alpha_{t+h-p})p(\alpha_{t+h-1}|\alpha_{t+h-2}, \cdots, \alpha_{t+h-1-p})p(\boldsymbol{\alpha}^{(t+h-2)}, y_t)}{p(\boldsymbol{\alpha}^{(t+h)})}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\prod_{i=1}^{h} p(\alpha_{t+i}|\alpha_{t+i-1}, \cdots, \alpha_{t+i-p}) p(\boldsymbol{\alpha}^{(t)}, y_t)}{p(\boldsymbol{\alpha}^{(t+h)})} \\
&= \frac{\prod_{i=1}^{h} p(\alpha_{t+i}|\alpha_{t+i-1}, \cdots, \alpha_{t+i-p}) p(y_t|\alpha_t) p(\boldsymbol{\alpha}^{(t)})}{p(\boldsymbol{\alpha}^{(t+h)})} \\
&= \frac{p(y_t|\alpha_t) \prod_{i=1}^{h} p(\alpha_{t+i}|\alpha_{t+i-1}, \cdots, \alpha_{t+i-p}) p(\boldsymbol{\alpha}^{(t)})}{\prod_{i=1}^{h} p(\alpha_{t+i}|\alpha_{t+i-1}, \cdots, \alpha_{t+i-p}) p(\boldsymbol{\alpha}^{(t)})} \\
&= p(y_t|\alpha_t)
\end{aligned}
$$

□

Proposition B.2. *Let $\{\alpha_t\}$ and $\{\tilde{\alpha}_t\}$ be two independent AR(p) and AR(q) processes, respectively. Assume that*

$$p(y_t|\alpha_t) = p(y_t|\boldsymbol{\alpha}^{(t)}, \tilde{\boldsymbol{\alpha}}^{(t)}), \qquad (t = 1, 2, \cdots)$$

Then

$$p(y_t|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)}) = p(y_t|\alpha_t), \qquad (h = 1, 2, \cdots)$$

Proof.

$$
\begin{aligned}
p(y_t|\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)}) &= \frac{p(y_t, \boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})}{p(\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})} \\
&= \frac{p(\alpha_{t+h}|\boldsymbol{\alpha}^{(t+h-1)}, \tilde{\boldsymbol{\alpha}}^{(t+h)}, y_t) p(y_t, \boldsymbol{\alpha}^{(t+h-1)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})}{p(\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})} \\
&= \frac{p(\alpha_{t+h}|\alpha_{t+h-1}, \cdots, \alpha_{t+h-p}) p(\tilde{\alpha}_{t+h}|\boldsymbol{\alpha}^{(t+h-1)}, \tilde{\boldsymbol{\alpha}}^{(t+h-1)}, y_t) p(y_t, \boldsymbol{\alpha}^{(t+h-1)}, \tilde{\boldsymbol{\alpha}}^{(t+h-1)})}{p(\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})} \\
&= \frac{p(\alpha_{t+h}|\alpha_{t+h-1}, \cdots, \alpha_{t+h-p}) p(\tilde{\alpha}_{t+h}|\tilde{\alpha}_{t+h-1}, \cdots, \tilde{\alpha}_{t+h-q}) p(y_t, \boldsymbol{\alpha}^{(t+h-1)}, \tilde{\boldsymbol{\alpha}}^{(t+h-1)})}{p(\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})} \\
&= \frac{\prod_{i=1}^{h} p(\alpha_{t+i}|\alpha_{t+i-1}, \cdots, \alpha_{t+i-p}) p(\tilde{\alpha}_{t+i}|\tilde{\alpha}_{t+i-1}, \cdots, \tilde{\alpha}_{t+i-q}) p(y_t, \boldsymbol{\alpha}^{(t)}, \tilde{\boldsymbol{\alpha}}^{(t)})}{p(\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})} \\
&= \frac{\prod_{i=1}^{h} p(\alpha_{t+i}|\alpha_{t+i-1}, \cdots, \alpha_{t+i-p}) p(\tilde{\alpha}_{t+i}|\tilde{\alpha}_{t+i-1}, \cdots, \tilde{\alpha}_{t+i-q}) p(y_t|\boldsymbol{\alpha}^{(t)}, \tilde{\boldsymbol{\alpha}}^{(t)}) p(\boldsymbol{\alpha}^{(t)}, \tilde{\boldsymbol{\alpha}}^{(t)})}{p(\boldsymbol{\alpha}^{(t+h)}, \tilde{\boldsymbol{\alpha}}^{(t+h)})} \\
&= \frac{p(y_t|\alpha_t) \prod_{i=1}^{h} p(\alpha_{t+i}|\alpha_{t+i-1}, \cdots, \alpha_{t+i-p}) p(\tilde{\alpha}_{t+i}|\tilde{\alpha}_{t+i-1}, \cdots, \tilde{\alpha}_{t+i-q}) p(\boldsymbol{\alpha}^{(t)}, \tilde{\boldsymbol{\alpha}}^{(t)})}{\prod_{i=1}^{h} p(\alpha_{t+i}|\alpha_{t+i-1}, \cdots, \alpha_{t+i-p}) p(\tilde{\alpha}_{t+i}|\tilde{\alpha}_{t+i-1}, \cdots, \tilde{\alpha}_{t+i-q}) p(\boldsymbol{\alpha}^{(t)}, \tilde{\boldsymbol{\alpha}}^{(t)})}
\end{aligned}
$$

$= p(y_t|\alpha_t)$. □

Bibliography

[1] M. A. Al-Osh and A. A. Alzaid. First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8:261–275, 1987.

[2] J. Albert. *Bayesian Computation with R*. Springer,New York, 2007.

[3] P. Berkhout. A bivariate Poisson count data model using conditional probabilities. *Statistica Neerlandica*, 58:349–364, 2004.

[4] D. Böhning. Zero-inflated Poisson models and C.A.MAN: A tutorial collection of evidence. *Biometrical Journal*, 40:833–843, 1998.

[5] J. P. Boucher, M. Denuit, and M. Guillén. Correlated random effects for hurdle models applied to claim counts. *Variance*, 5:68–81, 2011.

[6] K. Brännäs. Explanatory variables in the AR(1) count data model. *Umeå Economic Studies*, 381, 1995.

[7] P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer,New York, 2002.

[8] S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455, 1998.

[9] K. P. Burnham and D. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer,New York, 2002.

[10] A. C. Cameron and P. K. Trivedi. *Regression Analysis of Count Data*. Cambridge University Press, 2013.

[11] M. J. Campbell. Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature. *Journal of the Royal Statistical Society*, 157:191–208, 1994.

[12] C. Chatfield. *The Analysis of Time Series: An Introduction*. Chapman & Hall/CRC, 2004.

[13] J. S. Clark and A. E. Gelfand. A future for models and data in environmental science. *Trends in Ecology and Evolution*, 21:375–380, 2006.

[14] D. R. Cox. Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, 8:93–115, 1981.

[15] M. L. Dalrymple, I. L. Hudson, and R. P. K. Ford. Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Computational Statistics and Data Analysis*, 41:491–504, 2003.

[16] R. A. Davis. A negative binomial model for time series of counts. *Biometrika*, 96:735–749, 2009.

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

[18] R. A. Davis, W. T. M. Dunsmuir, and Y. Wang. On autocorrelation in a Poisson regression model. *Biometrika*, 87:1030–1039, 2000.

[19] B. Dennis and M. L. Taper. Density dependence in time series observations of natural populations: estimation and testing. *Ecological Monographs*, 64:205–224, 1994.

[20] M. J. Dobbie and A. H. Welsh. Modelling correlated zero-inflated count data. *Australian and New Zealand Journal of Statistics*, 43:431–444, 2001.

[21] K. El-Basyouny, S. Barua, and M. T. Islam. Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. *Accedent Analysis and Prevention*, 73:91–99, 2014.

[22] F. Famoye and K. P. Singh. Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*, 4:117–130, 2006.

[23] R. Ferland, A. Latour, and D. Oraichi. Integer-valued GARCH process. *Journal of Time Series Analysis*, 27:923–942, 2006.

[24] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Boca Raton: CRC Press, 2013.

[25] A. Gelman and D. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–511, 1992.

[26] C. J. Geyer. *Introduction to Markov Chain Monte Carlo, In Handbook of Markov Chain Monte Carlo edited by S. Brooks, A. Gelman, G. Jones and X. L. Meng.* CRC Press, London, 2011.

[27] S. K. Ghosh, P. Mukhopadhyay, and J. C. Lu. Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference*, 136:1360–1375, 2006.

[28] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Boca Raton, Fla: Chapman & Hall, 1998.

[29] S. Gurmu and J. Elder. Flexible bivariate count data regression models. *Journal of Business and Economics Statistics*, 30:265–274, 2012.

[30] D. B. Hall. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, 56:1030–1039, 2000.

[31] D. C. Heilbron. Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*, 36:531–547, 1994.

[32] J. M. Hilbe. *Negative Binomial Regression*. Cambridge University Press, New York, 2011.

[33] N. Johnson, S. Kotz, and N. Balakrishnan. *Multivariate Discrete Distributions*. Wiley, New York, 1997.

[34] B. Jϕrgensen, S. Lundbye-Christensen, P. Song, and L. Sun. A state space model for multivariate longitudinal count data. *Biometrika*, 86:169–181, 1999.

[35] M. R. Karim and S. L. Zeger. Generalized linear models with random effects; salamander mating revisited. *Biometrics*, 48:631–644, 1992.

[36] D. Karlis and L. Meligkotsidou. Multivariate Poisson regression with covariance structure. *Statistics and Computing*, 15:255–265, 2005.

[37] D. Karlis and P. Tsiamyrtzis. Exact Bayesian modeling for bivariate Poisson data and extensions. *Statistics and Computing*, 18:27–40, 2008.

[38] S. Kocherlakota and K. Kocherlakota. *Bivariate Discrete Distributions*. Marcel Dekker, Inc, 1992.

[39] S. Kocherlakota and K. Kocherlakota. Regression in the bivariate Poisson distribution. *Communications in Statistics-Theory and Methods*, 30:815–825, 2001.

[40] J. Lakshminarayana, S. Pandit, and K. S. Rao. On a bivariate Poisson distribution. *Communications in Statistics-Theory and Methods*, 28:267–276, 1999.

[41] D. Lambert. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34:1–14, 1992.

[42] S. R. Lele, B. Dennis, and F. Lutscher. Data cloning: easy maximum likelihood estimation for complex ecological models using bayesian markov chain monte carlo methods. *Ecology Letters*, 10:551—-563, 2007.

[43] S. R. Lele, K. Nadeem, and B. Schmuland. Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, 105:1617–1625, 2010.

[44] H. Liu. *Some Models for Time Series of Counts*. Ph.D. Thesis, Columbia University, 2012.

[45] P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall, London, 1989.

[46] C. E. McCulloch and S. R. Searle. *Generalized, Linear, and Mixed Models*. Hoboken, New Jersey: John Wiley & Sons, 2008.

[47] E. McKenzie. Some simple models for discrete variate time series. *Water Resources Bulletin*, 21:645—-650, 1985.

[48] Y. Min and A. Agresti. Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling*, 5:1–19, 2005.

[49] J. Mullahy. Specification and testing of some modified count data models. *Journal of Econometrics*, 33:341–365, 1986.

[50] K. B. Newman, S. T. Buckland, S. T. Lindley, L. Thomas, and C. Fernandez. Hidden process models for animal population dynamics. *Ecological Applications*, 16:74–86, 2006.

[51] I. Ntzoufras. *Bayesian Modeling Using WinBUGS*. John Wiley and Sons, Inc., Hoboken, New Jersey, 2009.

[52] X. Pedeli and D. Karlis. A bivariate INAR(1) process with application. *Statistical Modelling*, 11:325—-349, 2011.

[53] J. M. Ponciano, M. L. Taper, B. Dennis, and S. R. Lele. Hierarchical models in ecology: confidence intervals, hypothesis testing, and model selection using data cloning. *Ecological Society of America*, 90:356–362, 2009.

[54] M. D. Porter and G. White. Self-exciting hurdle models for terrorist activity. *The Annals of Applied Statistics*, 6:106–124, 2012.

[55] A. S. Quoreshi. Bivariate time series modeling of financial count data. *Communications in Statistics-Theory and Methods*, 35:1343–1358, 2006.

[56] D. V. Ravenzwaaij, P. Cassey, and S. D. Brown. A simple introduction to Markove Chain Monte-Carlo sampling. *Psychon Bulletin & Review*, 25:143–154, 2018.

[57] M. Ridout, J. Hinde, and C. G. B. Demétrio. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, 57:219–223, 2001.

[58] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.

[59] C. E. Rose, K. A. Martin, K. A. Wannemuehler, and B. D. Plikaytis. On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *Journal of Biopharmaceutical Statistics*, 16:463–481, 2006.

[60] P. Sólymos. dclone: data cloning in R . *The R Journal*, 2:29—-37, 2010.

[61] P. X. Song. *Correlated Data Analysis: Modeling, Analytics, and Applications*. Springer, New York, 2007.

[62] S. M. Stigler. *Statistics on The Table: The History of Statistical Concepts and Methods*. Harvard University Press, Cambridge, Mass., 2002.

[63] L. Tierney. Markov chains for exploring posterior distributions. *The Annalas of Statistics*, 22:1701–1762, 1994.

[64] E. G. Tsionas. Bayesian multivariate Poisson regression. *Communications in Statistics-Theory and Methods*, 30:243–255, 2001.

[65] P. D. Valpine. Better inferences from population-dynamics experiments using Monte Carlo state-space likelihood methods. *Ecological Society of America*, 84:3064–3077, 2003.

[66] A. M. Walker. On the asymptotic behavior of posterior distributions. *Journal of the Royal Statistical Society*, 31:80–88, 1969.

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

Vita Auctoris

Huda Al-Wahsh was born in 1973 in Kuwait. She obtained her B.Sc in Mathematics from the University of Jordan in 1996, then she worked as a teacher in the Ministry of Education in Jordan. She obtained her M.Sc. in Mathematics from the University of Windsor in 2014. She is currently a candidate for a Ph.D. in Statistics at the University of Windsor and will graduate in Summer 2018.