An Analysis of Claim Frequency and Claim Severity for Third Party Motor Insurance Using Monte Carlo Simulation Techniques

by

Cedric Dumais

A thesis submitted in partial fulfilment

of the requirements for the degree of

Master of Science (MSc) in Computational Sciences

The Faculty of Graduate Studies

Laurentian University

Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis Titre de la thèse	An Analysis of Claim Frequency and Claim Severity for Third Party Motor Insurance Using Monte Carlo Simulation Techniques
Name of Candidate Nom du candidat	Dumais, Cedric
Degree Diplôme	Master of Science
Department/Program Département/Programme	Computational Sciences
Date of Defence Date de la soutenance	August 22, 2019

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Peter Adamic
(Supervisor/Directeur(trice) de thèse)

Dr. Youssou Gningue
(Committee member/Membre du comité)

Dr. Hafida Boudjellaba
(Committee member/Membre du comité)

Dr. Franck Adékambi
(External Examiner/Examinateur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
Monsieur David Lesbarrères
Dean, Faculty of Graduate Studies
Doyen, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

I, **Cedric Dumais**, hereby grant to Laurentian University and/or its agents the non-exclusive license to archive and make accessible my thesis, dissertation, or project report in whole or in part in all forms of media, now or for the duration of my copyright ownership. I retain all other ownership rights to the copyright of the thesis, dissertation or project report. I also reserve the right to use in future works (such as articles or books) all or part of this thesis, dissertation, or project report. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that this copy is being made available in this form by the authority of the copyright owner solely for the purpose of private study and research and may not be copied or reproduced except as permitted by the copyright laws without written authority from the copyright owner.

Abstract

The purpose of this thesis is to introduce the reader to Multiple Regression and Monte Carlo simulation techniques in order to find the expected compensation cost the insurance company needs to pay due to claims made. With a fundamental understanding of probability theory, we can advance to Markov chain theory and Monte Carlo Markov Chains (MCMC). In the insurance field, in particular non-life insurance, expected compensation is very important to calculate the average cost of each claim. Applying Markov models, simulations will be run in order to predict claim frequency and claim severity. A variety of models will be implemented to compute claim frequency. These claim frequency results, along with the claim severity results, will then be used to compute an expected compensation for third party auto insurance claims. Multiple models are tested and compared.

Keywords

Regression, MCMC, Gibbs Sampler, Logistic, Poisson, Negative Binomial, Zero-Inflated, Insurance, Claim Frequency, Claim Severity, Expected Compensation

Contents

List of Tables

List of Figures

1 Introduction

In today's society, we place a high importance on modelling and predicting various types of risks. This allows for protection against various financial insecurities that might otherwise cause significant harm to our financial security. Such risks we model include non-life insurance or property and casualty insurance. In the field of personal casualty insurance, actuaries are often tasked with modelling auto insurance claims. The goal of the insurance company is to calculate an effective insurance price or premium to the corresponding insured party in order to cover the necessary risk. Claim frequency, also known as count data, and claim severity are the variables used to calculate the average cost of claims for property and casualty insurance. A superior model for claim frequency and claim severity means more competitive fees and in turn, a more profitable coverage for the insurer. Therefore, modelling claim frequency and claim severity is a crucial step for pricing personal and casualty insurance.

In the past, there has been extensive interest in count data models, particularly in Actuarial Science. Generalized linear models (GLMs) were given a life of their own by [Nelder and Wedderburn, 1972]. The use of generalized linear models by statisticians and actuaries has been discussed by [Haberman and Renshaw, 1996] as well as [Renshaw, 1994] and [McCullagh and Nelder, 1989]. Although these GLM models do many things well, they have several disadvantages. The assumptions made by GLM's may not hold true and therefore, the predictiveness of the model can be suboptimal. Residuals of the GLM's in insurance data are rarely homogeneous; an important feature in scoring a mod-

els fit. Lastly, GLM's only offer correlation between variables rather than causation. As shown in [Panjer and Willmot, 1983], the statistical interpretation of risk is essentially Bayesian. The approach adopted here is fully Bayesian, allowing for causation, model flexibility and credible intervals. To actuate this Bayesian approach, Markov Chain Monte Carlo (MCMC) is used for parameter estimation.

In this thesis statistical models for the claim frequency in Third Party Motor insurance are compared. Poisson regression has been the primary regression to model claim frequencies in the past. As has been shown in [Gourieroux and Jasiak, 2001], the Poisson distribution has many limitations due to its equidispersion. Equidispersion is defined as the equality of mean and variance within a distribution. To give an alternative, the Poisson regression model is compared with the Negative Binomial regression. Subsequently, the Zero-inflated models are compared as well. The claim severity is modeled by a Gamma model. In order to predict the expected compensation, the expected claim frequency is multiplied by the expected claim severity. The MCMC with Gibbs sampling will be used for parameter estimation. The models will then be scored by fit, and compared using the Deviance Information Criterion (DIC). A comparison of cost savings when using the best fitted model is given. Based on these models, the total claim severity can be simulated for premium calculation. The dependencies between the number of claims and claim severity is allowed. This regression takes into consideration the Poisson distribution as well as the Negative Binomial distribution to model the claim frequency and the Gamma distribution

to model the claim severity.

The contributions of this thesis are to provide four models to quantify and estimate claim frequency. Provide a Gamma model to estimate claim severity. Compute the expected compensation for the different zones of policyholders. Discuss strengths and weaknesses of each count variable model obtained by MCMC. Discuss and provide an example of regression model comparison with a Bayesian approach for auto claim frequencies.

2 Literature Review

In this chapter the review of the literature employed in the thesis is explained. An introduction to Poisson regression, Negative Binomial regression and Gamma regression is given. Markov Chain Monte Carlo (MCMC) and Gibbs sampling is reviewed, along with the Deviance Information Criterion used to compare the models.

2.1 Poisson Model

To begin, the Poisson distribution is introduced. The benchmark model for count data is the Poisson distribution. It is useful at the outset to review some fundamental properties and characterization results. If the discrete random variable Y is Poisson-distributed with intensity or rate parameter μ, $\mu > 0$, and t is the exposure, defined as the length of time

during which the events are recorded, then Y has density of the Poisson distribution

$$Pr[Y = y] = \frac{e^{-\mu t}(\mu t)^y}{y!}, \qquad y = 0, 1, 2... \tag{1}$$

where $E[Y] = V[Y] = \mu t$. If the length of the exposure period t is equal to unity, then

$$Pr[Y = y] = \frac{e^{-\mu}\mu^y}{y!}, \qquad y = 0, 1, 2, ... \tag{2}$$

This distribution has a single parameter μ, and is referred to as $P[\mu]$. Its k^{th} raw moment, $E[Y^k]$, may be derived by differentiating the moment generating function k times. The Poisson distribution has equal mean and variance. This is referred as equidispersion. This property is often violated in real-life data. Overdispersion (underdispersion) means the variance exceeds (is less than) the mean.

The law of rare events states that the total number of events will follow, approximately, the Poisson distribution if an event may occur in any of a large number of trials but the probability of occurrence in any given trial is small. More formally, let $Y_{n,\pi}$ denote the total number of successes in a large number n of independent Bernoulli trials with success probability π of each trial being small. Then

$$Pr[Y_{n,\pi} = k] = \binom{n}{k}\pi^k(1-\pi)^{n-k}, \qquad k = 0, 1, ..., n. \tag{3}$$

2.1.1 Poisson Regression Model

A standard application of Poisson regression is to cross-section data [Cameron and Trivedi, 2013]. Typical cross-section data for applied work consist of n independent observations, the i^th of which is (y_i, x_i). The scalar dependent variable, y_i , is the number of

occurrences of the event of interest, and x_i is the vector of linearly independent regressors that are thought to determine y_i . A regression model based on this distribution follows by conditioning the distribution of y_i on a k-dimensional vector of covariates, $x_i = [x_{1i}, ..., x_{ki}]$, and parameters β, through a continuous function $\mu(x_i, \beta)$, such that $E[y_i|x_i] = \mu(x_i, \beta)$. That is, y_i given x_i is Poisson-distributed with probability density

$$f(y_i|x_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \qquad y_i = 0, 1, 2, \ldots \tag{4}$$

In the log-linear version of the model, the mean parameter is parameterized as

$$\mu_i = exp(x_i'\beta), \tag{5}$$

to ensure $\mu > 0$. These two equations jointly define the Poisson (log-linear) regression model. For notational economy $f(y_i|x_i)$ is written in place of the more formal $f(Y_i = y_i|x_i)$, which distinguishes between the random variable Y and its realization y. By the property of the Poisson, $V[y_i|x_i] = E[y_i|x_i]$, implying that the conditional variance is not a constant, and hence the regression is intrinsically heteroskedastic. This is to say that the variance of the residual terms varies widely. In the log-linear version of the model the mean parameter is parameterized as Equation (5), which implies that the conditional mean has a multiplicative form given by

$$\begin{aligned} E(y_i|x_i) &= exp(x_i'\beta) \\ &= exp(x_{1i}\beta_0)exp(x_{2i}\beta_1)\cdots exp(x_{ki}\beta_k), \end{aligned} \tag{6}$$

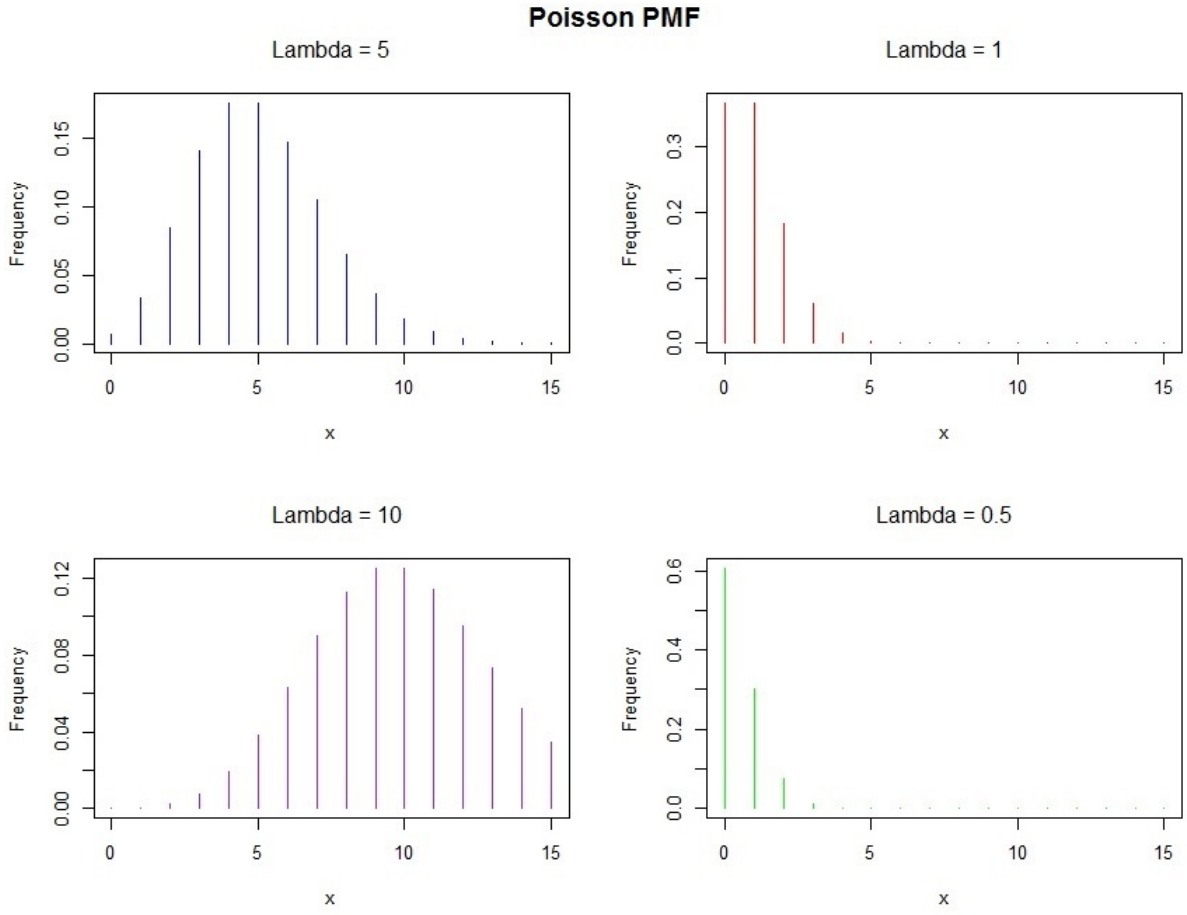
with interest often lying in changes in this conditional mean due to changes in the regressors. The additive specification, $E[y_i|x_i] = x_i'\beta = \sum_{j=1}^{k} x_{ji}\beta_i$, is likely to be unsatisfactory

because certain combinations of β_i and x_i will violate the nonnegativity restriction on μ_i. In most cases, $x_1 = 1$, β_0 is interpreted as the intercept, and $\beta_1, \ldots, \beta_k$ are unknown parameters. For example, let y be the observed number of accidents, N be the known exposure risk, and x the known explanatory variables. The mean number of events μ is expressed as the product of N and t_i. This is often called the rate of occurrence. That is, $\mu(x) = N(x)t_i(x, \beta)$. The expected value for this rate of occurrence with exposure variable t_i is

$$\begin{aligned} E(y_i|x_i) &= t_i\lambda(x_i'\beta) \\ &= t_i exp(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki}). \end{aligned} \tag{7}$$

Therefore, for a given set of regressor variables, the outcome (dependent variable) follows a Poisson Distribution.

Figure 2.1: Probability Mass function for Poisson Distributions



2.2 Negative Binomial Model

In this thesis the commonly known NB2 model introduced by [Cameron and Trivedi, 1986] is used. The most common way this model is derived is by assuming the data are Poisson, but there is gamma-distributed unobserved individual heterogeneity reflecting the fact that the true mean is not perfectly observed. This model has a scale parameter of $\alpha = \frac{1}{v}$.

$$Pr(Y = y_i|\mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\alpha^{-1} + \mu_i}\right)^{y_i}, \tag{8}$$

where

$$\mu_i = t_i\mu$$

$$\alpha = \frac{1}{v}.$$

The parameter μ is the mean of y per unit of exposure t_i. The mean is calculated by

$$E(y_i) = \mu_i = e^{x_i^t\beta}, \tag{9}$$

and the variance by

$$Var(y_i) = \sigma_i^2 = \mu_i(1 + \alpha\mu_i). \tag{10}$$

It is shown below that in Equation (8) when $\alpha \to 0$, the Poisson distribution is obtained. This is because α is known as the dispersion parameter. α^{-1} is referred to as the index

or dispersion parameter. From [Cameron and Trivedi, 2013],

$$\begin{aligned} f(y) &= \left(\prod_{j=0}^{y=1}(j+\alpha)\right)\frac{1}{y!}\left(\frac{\alpha}{\alpha+\mu}\right)^{\alpha}\left(\frac{1}{\alpha+\mu}\right)^{y}\mu^{y} \\ &= \left(\prod_{j=0}^{y=1}\frac{j+\alpha}{\alpha+\mu}\right)\left(\frac{\alpha}{\alpha+\mu}\right)^{\alpha}\mu^{y}\frac{1}{y!} \\ &= \left(\prod_{j=0}^{y=1}\frac{1+\frac{j}{\alpha}}{1+\frac{\mu}{\alpha}}\right)\left(\frac{1}{1+\frac{\mu}{\alpha}}\right)^{\alpha}\mu^{y}\frac{1}{y!} \\ &\to 1e^{-\mu}\mu^{y}\frac{1}{y!} \qquad as \quad \alpha\to\infty. \end{aligned}$$

Since $1e^{-\mu}\mu^{y}\frac{1}{y!}$ is the Poisson distribution, the Poisson is the special case as $\alpha \to 0$. As with the Poisson distribution, the parameter μ is the incidence rate per unit of exposure t_i. The function $\Gamma(\cdot)$ is the gamma function and defined by

$$\Gamma(\alpha) = \int_0^{\infty} e^{-t}t^{\alpha-1}dt, \qquad \alpha > 0. \tag{11}$$

Therefore, the Negative Binomial distribution given by Equation (8) is represented as a Poisson-Gamma mixture distribution.

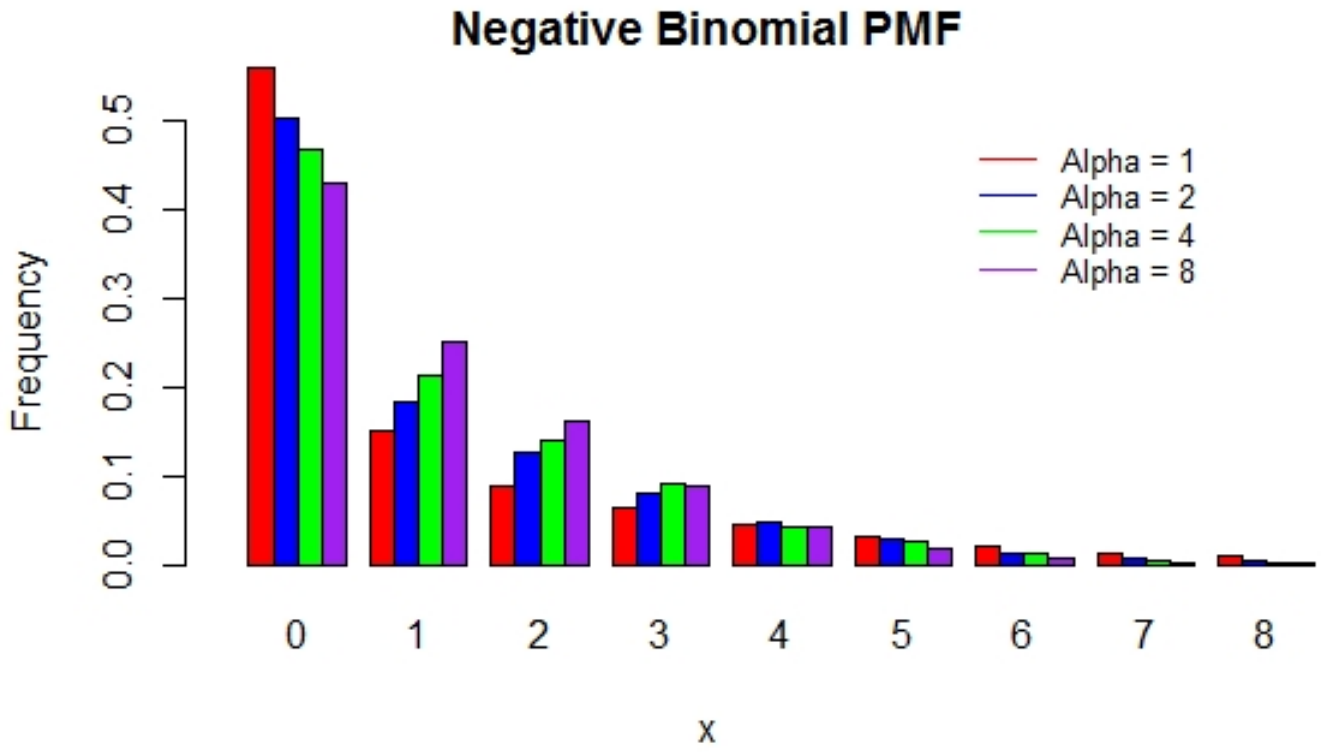
2.2.1 Negative Binomial Regression Model

The Negative Binomial regression model is determined by exposure time t and k regressor variables. The expected value is expressed as

$$\mu_i = exp(ln(t_i) + \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}). \tag{12}$$

In Negative Binomial regression, where μ is the mean and β_0 is the intercept. The regression coefficients $\beta_1, \beta_2, \ldots, \beta_k$ are unknown parameters that are estimated from the data

Figure 2.2: Probability Mass function for Negative Binomial Distributions with $\mu = 2$.



set. Therefore, a Negative Binomial regression model for an observation i is given as

$$Pr(Y = y|\mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\alpha^{-1}} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i}, \tag{13}$$

where y is the dependent variable.

2.3 Zero-Inflated Poisson Model

The zero-inflated Poisson model was introduced by [Lambert, 1992] to solve problems with the large zero counts in data. The zero-inflated models are a solution to incorrect

and biased models, incorrect parameter estimations, biased errors and over-dispersion caused by the high zero counts in the dataset. The zero-inflated Poisson is a two part model, consisting of both binary and count model sections. Suppose that for each observation i, there are two possible cases. If case one occurs, the count is zero. If case two occurs, counts(including zeros) are generated according to the Poisson model. Given the probability of case one to be π, and the probability of case two to be $1-\pi$, the probability distribution of the Poisson is

$$\begin{aligned} Pr(y_i = 0) &= \pi_i + (1-\pi_i)e^{-\mu_i} \\ Pr(y_i > 0) &= (1-\pi_i)\frac{\mu_i^{y_i} exp(-\mu_i)}{y_i!}, \end{aligned} \tag{14}$$

and,

$$\begin{aligned} Var(y_i) &= (1-\pi_i)(\mu_i + \mu_i^2) \\ &> \mu_i(1-\pi_i) = E(y_i). \end{aligned}$$

The zero-inflated Poisson can be written as

$$\mu_i = exp(ln(t_i) + \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}),$$

where t_i is the exposure variable and β_i are the regressor variables. The regressor coefficients can then be estimated using maximum likelihood estimation or in this case, Bayesian estimation. The goal is to estimate (β, γ). [Lambert, 1992] introduced the model in which $\mu_i = \mu(x_i, \beta)$ and the probability π_i is parameterized as a logistic function of the

observable vector of covariates z_i. The logistic link function is given as follows:

$$\begin{aligned} y_i &= 0, \qquad \text{with probability} \quad \pi_i \\ y_i &\sim P(\mu_i), \qquad \text{with probability} \quad (1-\pi_i) \\ \pi_i &= \frac{exp(z_i'\gamma)}{1+exp(z_i'\gamma)}, \end{aligned} \tag{15}$$

where

$$\exp(z_i'\gamma) = exp(ln(t_i) + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \ldots + \gamma_m z_{mi}). \tag{16}$$

2.4 Zero-Inflated Negative Binomial Model

The zero-inflated Negative Binomial (ZINB) regression model is similar to the Poisson zero-inflated model. A zero-inflated Negative Binomial model will enable us to distinguish between the effect of the splitting mechanism and the over-dispersion induced by individual heterogeneity [Greene, 1994]. Suppose that for each observation i, there are two possible cases. If case one occurs, the count is zero. If case two occurs, counts(including zeros) are generated according to the Negative Binomial model. Given the probability of case one to be π, and the probability of case two to be $1-\pi$, the probability distribution of the ZINB is

$$\begin{aligned} Pr(y_i = 0) &= \pi_i + (1-\pi_i)g(y_i = 0) \quad \text{if } i = 0 \\ Pr(y_i = k) &= (1-\pi_i)g(y_i) \quad \text{if } i > 0, \end{aligned} \tag{17}$$

where π_i is the logistic link function defined by

$$\exp(z_i'\gamma) = exp(ln(t_i) + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \ldots + \gamma_m z_{mi}). \tag{18}$$

The zero-inflated Negative Binomial model can also be expressed as

$$\mu_i = exp(ln(t_i) + \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki})$$

where t_i is the exposure variable and β_i are the regressor variables. The regressor coefficients are then estimated using Bayesian estimation.

2.5 Gamma Model

The Gamma distribution is a two-parameter family of continuous probability distributions. It has parameter β and shape parameter α, $\alpha > 0$ and $\beta > 0$. As shown in [Gelman et al., 2013], the Gamma distribution is the conjugate prior distribution for the inverse of the normal variance and for the mean parameter of the Poisson distribution. Let θ be a random variable and $\theta \sim Gamma(\alpha, \beta)$; that is, the random variable θ follows a Gamma distribution with parameters α and β. The density function of a Gamma distribution given by the random variable θ is

$$f(\theta; \beta, \alpha) = \frac{\beta}{\Gamma(\alpha)} (\beta\theta)^{\alpha-1} e^{-\beta\theta} I_{(0,\infty)}(\theta), \tag{19}$$

where $I(.)$ is an indicator function. Under this parameterization, the mean is given by

$$E(\theta) = \frac{\alpha}{\beta}, \tag{20}$$

and the variance by

$$\begin{aligned} Var(\theta) &= \frac{\alpha}{\beta^2} \\ &= \frac{\mu^2}{\alpha}. \end{aligned} \tag{21}$$

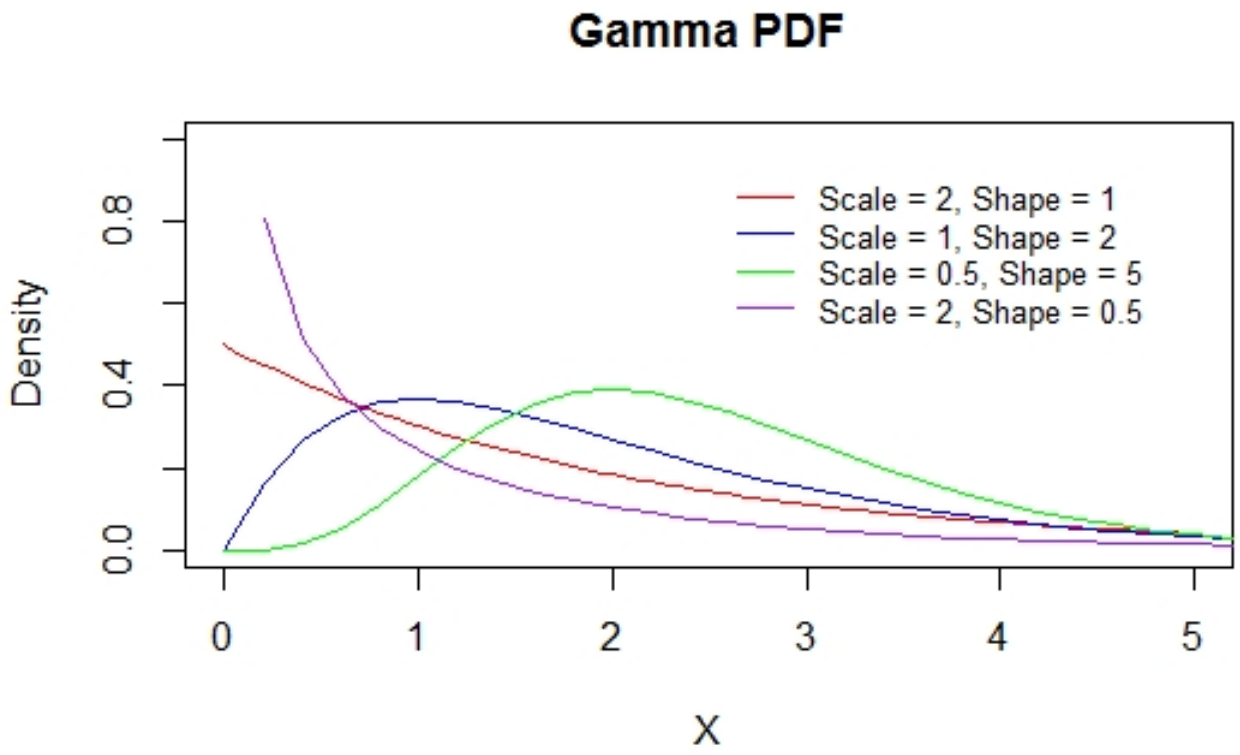


Figure 2.3: Density function for Gamma Distributions

By setting $\beta = \frac{\alpha}{\mu}$, it is shown in [Gelfand et al., 2005] that the Gamma density function can be written as

$$f(\theta) = \frac{1}{\theta\Gamma(\alpha)}\left(\frac{\alpha\theta}{\mu}\right)^{\alpha} exp\left(\frac{\alpha\theta}{\mu}\right) I_{(0,\infty)}(\theta). \tag{22}$$

2.5.1 Gamma Regression Model

Let $Y_i \sim G(\mu_i, \alpha)$ for $i = 1, 2, \ldots, n$, be independent random variables. Then the Gamma regression model is defined as

$$\begin{aligned} g(\mu_i) &= x_i'\eta \\ \mu_i &= \eta_0 x_{0i} + \eta_1 x_{1i} + \ldots + \eta_k x_{ki}, \end{aligned} \tag{23}$$

where $\eta = (\eta_0, \eta_1, \ldots, \eta_p)'$ is a vector of unknown regression parameters $(p < n)$, $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ is the vector of p covariates and η_i is a linear predictor. As in most regression models, $x_{01} = 1$ for all i so that the model has intercept η_0.

2.6 Collective Risk Model

In the basic insurance risk model from [Embrechts et al., 2013], the number of claims and the total claim produced in a given time period $t = 1, 2, \ldots, T$ for some class i is denoted by (N_{it}, X_{it}) where

$$X_{it} = \begin{cases} \sum_{k=1}^{N_{it}} W_{itk} & N_{it} > 0 \\ 0 & \text{otherwise}, \end{cases} \tag{24}$$

and where W_{itk} is the amount of the kth claim at time t for some class i. The assumptions for the model are given in [Migon and Moura, 2005], and they are:

- The number of claims in the interval $(t-1, t]$ is a random variable denoted as N_{it}
- Conditional on $N_{it} = n_{it}$, the claim severity $W_{ik}, k = 1, 2, \ldots, n_{it}$, are positive independent and identically distributed random variables with finite mean $\mu_i = E(W_{ik})$ and variance $\sigma_i^2 = Var(W_{ik}) < \infty$.

- The claims occur at random times $t_{1i} \leq t_{2i} \leq \ldots$ and the inter-arrival times $T_{ji} = t_{ji} - t_{j-1,i}$ are assumed to be independent and identically exponentially distributed random variables with finite mean $E(T_{ji}) = \lambda_i^{-1}$.

By assuming the sequences T_j and W_j are mutually independent from each other and identically distributed, the above conditions hold. It follows that N_{it} is a homogeneous Poisson process with rate λ_{it}, then $\sum_{k=1}^{\pi_{it}} N_{itk}|\lambda_{it} \sim Poisson(\lambda_{it}\pi_{it})$, η_{it} is the observed number of claims at time t, for class i and π_{it} is the insured population at time t for class i, and not π_t, t. Assuming that $W_{itk} \sim Gamma(\alpha_{it}, \theta_{it})$, the inter-arrival times are exponentially distributed. The Poisson-Gamma model is given by

$$\begin{aligned} N_{it}|\lambda_{it}, \pi_{it} &\sim Poisson(\lambda_{it}\pi_{it}), \qquad \lambda_i > 0 \\ W_{itk} &\sim Gamma(\alpha_{it}, \theta_{it}), \\ X_{it}|\eta_{it}, \theta_{it} &\sim Gamma(\eta_{it}\alpha_{it}, \theta_{it}) \qquad \theta_i > 0 \end{aligned} \tag{25}$$

where $\alpha_{it} = \eta_{it}\alpha_i, \eta_{it}$ is the observed number of claims at time t, for class i and $\pi_i t$ is the insured population at time t for class i.

A similar idea is applied to the Negative Binomial expected values and the Gamma expected values. [Kaas et al., 2008] $S = X_1, X_2, \ldots, X_N$ and X_i are amounts for a claim and N is the total amount of claims and S is the sum of the collective claims. The expected value is given by

$$E(S) = E(N)E(X) \tag{26}$$

and the variance

$$Var(S) = E(N)Var(X) + (E(X))^2 Var(N). \tag{27}$$

Therefore, the premium for collective risk is given by

$$P = E(N)E(X). \tag{28}$$

2.7 Markov Chain Monte Carlo

MCMC is essentially Monte Carlo integration using Markov chains. Scientists need to integrate over possibly high-dimensional probability distributions to make inference about model parameters or to make predictions. Markov Chain Monte Carlo (MCMC) algorithms have made a significant impact on problems where Bayesian analyses can be applied; see [Spiegelhalter et al., 1996]. MCMC can be broken down into key steps, first, randomly generating numbers also known as the Monte Carlo part. Second, allow the numbers generated to influence the next number, also known as the Markov chain part. Third, check for convergence to a reasonable distribution. Monte Carlo integration evaluates $E[f(x)]$ by drawing samples $X_t, t = 1, \ldots, n$ from $\pi(.)$ and then approximating

$$E[f(X)] \approx \frac{1}{n} \sum_{t=1}^{n} f(X_t). \tag{29}$$

So the population mean of $f(X)$ is estimated by a sample mean. When the samples X_t are independent, the law of large numbers ensures that the approximation can be made as accurate as desired by increasing the sample size n. In general, drawing samples X_t independently from $\pi(.)$ is not feasible since $\pi(.)$ can be non-standard full conditional distributions. However, X_t does not necessarily need to be independent. It can be generated

by any process which draws samples throughout support of $\pi(.)$. Suppose $X_0, X_l, X_2, \ldots$ are generated as a sequence of random variables, such that at each time $t \sim 0$, the next state X_{t+l} is sampled from a distribution $P(X_{t+1}|X_t)$ which depends only on the current state of the chain, X_t. That is, given X_t, the next state X_{t+1} does not depend further on the history of the chain $(X_0, X_1, \ldots, X_{t-1})$. This sequence is called a Markov chain. In general, MCMC involves simulating from a complex and multivariate target distribution, $p(X)$, by generating a Markov chain with the target density as its stationary density, [Gelman et al., 2013]. Markov chain simulation is used when it is not possible or not computationally feasible to sample X directly from $p(X|y)$. Instead, the method draws iteratively in such a way that at each step, the draw from the distribution is expected to be closer to $p(X|y)$. The basic principle is that once the chain has run sufficiently long enough, it will approximate the posterior distribution $p(X|y)$. In general, $m \geq 1$ independent sequences of simulations are run, each with a length of n, $(X_{j1}, X_{j2}, \ldots, X_{jn})$ for $j = 1, \ldots, m$.

The term Markov chain stands for a sequence of random variables $X_1, X_2, \ldots$ for which, for any t, the distribution X_t depends only on the most recent variable, X_{t-1}. To begin an MCMC simulation is made by selecting a X_0 and then for each t, X_t is drawn from a transition distribution $T_t(X_t|X_{t-1})$ so that the Markov chain hopefully converges to the posterior distribution $p(X|y)$. Once algorithms have been implemented and the simulations drawn, it is extremely important to ensure the convergence of the simulated sequences. The sequence is monitored with a time-series plot, and the Gelman Rubin method is used to

diagnose convergence, [Brooks and Gelman, 1998]. More on how the MCMC is applied is discussed in Section 2.7.1.

The two most prevailing techniques used in MCMC are the Metropolis-Hastings algorithm and the Gibbs sampler. Bayes' theorem, can be conceptualized as

$$\textbf{posterior} \propto \textbf{prior} \times \textbf{likelihood}. \tag{30}$$

That is, the posterior is proportional to the likelihood times the prior. The guidelines below flow directly from this theorem:

- If the prior is uninformative, the posterior is determined by the data
- If the prior is informative, the posterior is a mixture of the prior and the data
- The more informative the prior, the more data is needed to influence the beliefs since the posterior is determined more so from the prior information
- If the dataset is large, the data will dominate the posterior distribution

2.7.1 Gibbs Sampler

The Gibbs sampler was introduced by [Geman and Geman, 1984]. Although the Metropolis-Hastings is more commonly used in the literature, the Gibbs sampler was implemented in this thesis since the data analysed is low-dimensional. Low-dimensional refers to the features of the dataset or the amount of variables in the model. High-dimensional data may consist of hundreds or thousands of features. Gibbs sampling can be understood as

running a sequence of low-dimensional conditional simulations. It is used when decomposition's into such conditionals are easy to implement and fast to run, which is the case in this thesis. As described in [Gill, 2002], the Gibbs sampler is useful in producing Markov chain values. It is a special case of the Metropolis-Hastings algorithm with a probability of acceptance of one. Suppose a joint density $f(x, y_1, \ldots, y_p)$ is given and the marginal density $f(x)$ is needed to calculate the marginal mean or variance. The most natural way to do so would be to integrate $f(x)$ directly. However, in some cases, it is simpler to sample from a conditional distribution than to marginalize by integrating over a joint distribution. Gibbs sampling generates a sample $X_1, \ldots, X_n \sim f(x)$ without requiring $f(x)$ [Casella and George, 1992]. By simulating a large enough sample, the mean, variance, or any other characteristic of $f(x)$ can be calculated. As an example, to calculate the mean of $f(x)$,

$$\lim_{m\to\infty} \frac{\sum_{i=1}^{m} X_i}{m} = \int_{-\infty}^{\infty} x f(x) dx = E[X]. \tag{31}$$

is used. By taking m large enough in Equation (31), any population characteristic, even the density itself, can be obtained to any degree of accuracy. The basic tenet of Gibbs sampling is that one can express each parameter to be estimated as conditional on all the others. By going through these conditional distributions, eventually the chain converges to the true joint distribution of interest. Suppose k samples are needed of $X = (x_1, \ldots, x_n)$ from a joint distribution $p(x_1^{(i)}, \ldots, x_n^{(i)})$. Let the i^{th} sample be denoted by $X^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$. Beginning with $X^{(i)}$, the objective is to obtain the next sample. This

sample $X^{(i+1)}$ is given by

$$X^{(i+1)} = \big(x_1^{(i+1)}, x_2^{(i+1)}, \ldots, x_n^{(i+1)}\big),$$

is a vector. Each component of the vector $x_j^{(i+1)}$ is sampled from the distribution of that component conditioned on all other components sampled so far. Therefore,

$$X_j^{i+1} \sim p\big(x_j^{(i+1)} | x_1^{(i+1)}, \ldots, x_{j-1}^{(i+1)}, x_{j+1}^{(i)}, \ldots, x_n^{(i)}\big)$$

is the $(i+1)^{th}$ component for the variable x_j. Notice that the i^{th} components of the $j+1$ variables are used. This is because the $(i+1)^{th}$ component has not been calculated yet. The above steps are repeated k times. From these steps, the expected value of any variable can be approximated by averaging over all the samples. Since the average over all the samples is used to calculate the characteristics of the variable, it is common to ignore a number of the samples at the beginning (often referred to as the burn-in period). The sample approximates the joint distribution of all variables since X^{i+1} approaches $p(X)$ as $i \to \infty$.

To help better describe what is going on with the Gibbs sampler, a simple example is explained. Suppose X and Y are two binary random variables with joint distribution $P(X = x, Y = y) = p_{X,Y}(x, y)$ given by the following table:

$$\begin{bmatrix} X \backslash Y & 0 & 1 \\ 0 & 0.6 & 0.1 \\ 1 & 0.15 & 0.15 \end{bmatrix}$$

That is, $p_{X,Y}(0, 0) = 0.6$. The conditional distribution of X is easily calculated from Bayes formula $P(A|B) = P(A \cap B)/P(B)$. For example, $P(X = 0|Y = 0) = P(X = 0 \cap Y =$

$0)/P(Y = 0) = 0.6/0.75 = 0.8$. Starting from some value of X, Y and proceeding to iterate the following two steps will achieve Gibbs sampling. Simulate a new value of X from $P(X|Y = y)$ where y is the current value of Y. Simulate a new value of Y from $P(Y|X = x)$ where x is the current value of X (generated in 1). Running this simulation via computer program, the summary of the first $n = 50$ iterations are kept. It is found that the proportion of the iterations in which $X = x$ and $Y = y$ is increasingly close to $P(X = x, Y = y) = p_{X,Y}(x, y)$. This is a result of simulating a Markov chain whose stationary distribution is $P(X = x, Y = y) = p_{X,Y}(x, y)$.

Once convergence is achieved, the simulated values are sampled from a distribution that asymptotically follows the target posterior distribution. By increasing the length of the chain (increasing n), the sampling variance of the posterior variables is decreased. The mean and standard deviation, as well as, the naïve standard error and time-series standard error are computed. These error values are measures of the computational MCMC error for the estimation of the posterior expected value. The naïve SE is given as

$$SE_n = \sqrt{\frac{Var(X)}{C \cdot S}},$$

where X is the vector of posterior samples, C is the number of chains run and S is the number of iterations run. Similarly, the time-series SE is given as

$$SE_{ts} = \sqrt{\frac{Var_{ts}(X)}{C \cdot S}},$$

where $Var_{ts}(X)$ is the average of the variance of each set of samples X for each chain C. In short, the naïve SE disregards autocorrelation where as the time-series SE takes into account the often high auto correlations found in MCMC sampling. The ensuing trace

plots are time-series plots of the sampled values [Toft et al., 2007]. They are the first tool used to assess the convergence of the chain. If the chain has reached convergence, then the time-series should be centered around a constant mean. If multiple chains with different starting points are plotted, then these plots should seem indistinguishable.

A second test for convergence of chains is the autocorrelation of the monitored parameters within each chain. High autocorrelations suggest slow mixing of chains and, usually, slow convergence to the posterior distribution [Smith et al., 2007]. Although the model converges eventually, they are significantly less optimal on computing time compared models with low autocorrelations which tend to converge much faster. A common strategy is to thin the chains to reduce sample autocorrelation. A chain can be thinned by keeping every k^{th} simulated draw from each sequence. Thinning the chain is an option considered to improve accuracy, however according to [Link and Eaton, 2012], for approximations of simple features of the target distribution (e.g. means, variances and percentiles), thinning is neither necessary nor desirable.

One approach to get an estimate of the severity of the variance is to run several chains and use the between-chains variance in $\hat{\theta}$. Specifically, if $\hat{\theta}_j$ denotes the estimate for chain $j(1 \geq j \geq m)$ where each the m chains have the same length, then an estimate for the variance is

$$Var(\hat{\theta}) = \frac{1}{n-1} \sum_{j=1}^{m} \left(\hat{\theta}_j - \hat{\theta*} \right)^2 \quad \text{where} \quad \hat{\theta*} = \frac{1}{n} \sum_{j=1}^{m} \hat{\theta}_j. \tag{32}$$

2.7.2 Heidelberg and Welch Test

A third test for convergence of the MCMC chain is the Heidelberg and Welch test [Heidelberger and Welch, 1983]. This is a two part test. In the first part, a test statistic on the entire chain is calculated. The null hypothesis is that the chain is from a stationary distribution. If the null hypothesis is rejected, the first 10% of the chain is discarded. This step is repeated until the null hypothesis fails to be rejected or 50% of the chain is discarded. For the second part, if the chain passes the first part of the diagnosis, it then takes the part of the chain not discarded from the first part to test the second part. The half-width test calculates half the width of the $(1-\alpha)\%$ credible interval around the mean. If the ratio of the half width and the mean is lower than some ϵ, then the chain passes the test. Otherwise, the chain must be run for more iterations. The test also provides a key p-value statistic for each variable. The null hypothesis for the test is that the Markov chain are from a stationary distribution.The stationary distribution of a Markov Chain with transition matrix P is some vector, X, such that $XP = X$. In other words, over the long run, no matter what the starting state was, the proportion of time the chain spends in state j is approximately X_j for all j. Simple put, the probability to end (converge) a very long random walk (Markov chain), is independent of where the walk started. A large p-value means we cannot reject the null hypothesis and that the Markov chains do have a stationary distribution. Therefore, large p-value among these tabled results are an indication for convergence and stationary distribution. This is to say, as $i \to \infty$ there is a unique solution.

2.8 Model Comparison

The deviance information criterion (DIC), introduced by [Spiegelhalter et al., 2002], for a probability model $p(y|\theta)$ with observed data $y = (y_1, \ldots, y_n)$ and unknown parameters θ is defined as

$$DIC = E[D(\theta|y)] + p_D. \tag{33}$$

It considers both model fit as well as model complexity. The goodness-of-fit is measured by the posterior mean of the Bayesian deviance $D(\theta)$ defined as

$$D(\theta) = -2logp(y|\theta) + 2logf(y), \tag{34}$$

where $f(y)$ is some fully specified standardizing term. Model complexity is measured by the number of parameters p_D defined by

$$p_D = E[D(\theta|y)] - D(E[\theta|y]). \tag{35}$$

The DIC criterion has been suggested as a criterion of model fit by [Spiegelhalter et al., 2002]. DIC is just the sum of the posterior mean deviance and the effective number of parameters. The goal is to find a model that will be best for prediction when taking into account the uncertainty inherent in sampling. According to the criterion, as with other information criterion, the model with the smallest DIC is preferred. The DIC does not give belief for a "true" model, instead, it provides model comparison for short term predictions. Using the MCMC output, both p_D and DIC are easily computed by taking the posterior mean of the deviance $E[D(\theta|y)]$ and the estimate of the deviance $D(E[\theta|y])$. Since $2logf(y)$ is a standardizing term that is a function of data alone, in the application

of this thesis the standardized term $2logf(y)$ from Equation (34) is set to zero by setting $f(y) = 1$.

3 Data

The dataset used in this thesis is a Third Party Motor Insurance dataset from Sweden. In Sweden, all motor insurance companies apply identical risk arguments to classify customers, and thus their portfolios and their claims statistics can be combined. The data was compiled by the Swedish Committee on the Analysis of Risk Premium in Motor Insurance. This dataset is used to apply the different multiple regressions to compare fits, scores, and show how valuable information can be extracted from the data. The data has a sample size of n=1937 with 7 variables. The variables are:

- **Kilometres:** Number of kilometres driven per year
- **Zone:** Geographical location divided by major cities and surroundings.
- **Bonus:** No claims made bonus. Equal to the number of years $+1$ since last claim
- **Make of Vehicle:** Represents the 8 most driven car models
- **Insured:** Amount of time insured in policy-years
- **Claims** Number of Claims made
- **Payment** Value amount of payment made in Swedish Krona (Skr)

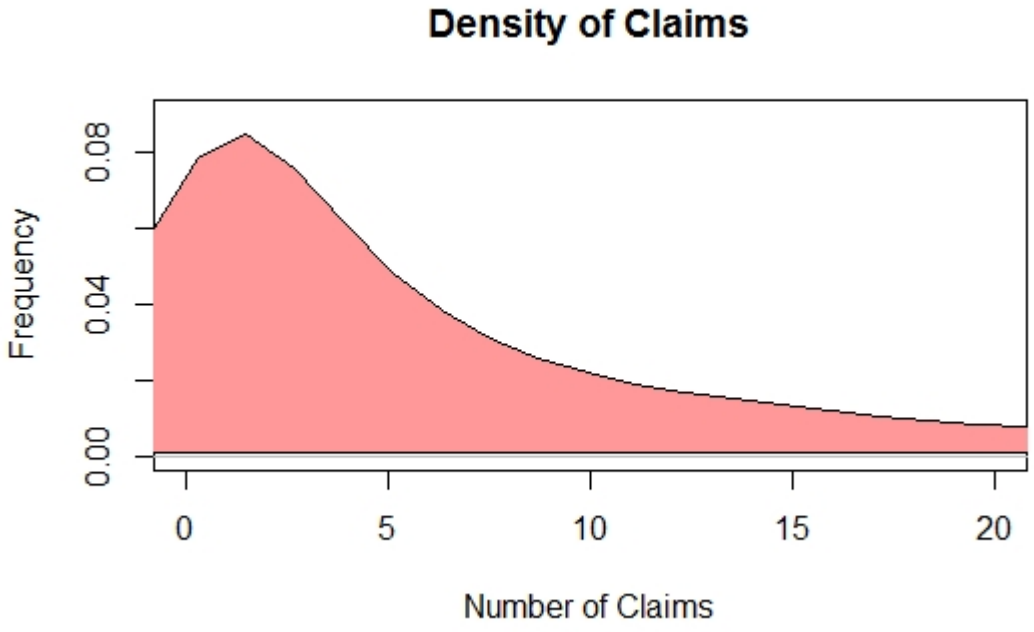
Table 3.1: Summary Statistics for the Third Party Insurance Data

	mean	sd	median	min	max
Kilometres	2.985793	1.410409	3.000	1.00	5.0
Zone	3.970211	1.988858	4.000	1.00	7.0
Bonus	4.015124	2.000516	4.000	1.00	7.0
Make	4.991751	2.586943	5.000	1.00	9.0
Insured	1092.195270	5661.156245	81.525	0.01	127687.3
Claims	51.865720	201.710694	5.000	0.00	3338.0
Payment	257007.644821	1017282.585648	27403.500	0.00	18245026.0

The Kilometres driven per year are separated into 5 different categories, $<$ 1000, 1000-15000, 15000-20000, 20000-25000, and $>$25000. The Zones are separated into 7 different zones, Stockholm, Goteborg, Malmo with surroundings, Other large cities with surroundings, Smaller cities with surroundings in southern Sweden, Rural areas in southern Sweden, Smaller cities with surroundings in northern Sweden, Rural areas in northern Sweden, and Gotland (island). Unfortunately, the make of the cars was never released due to the potential impact on sales. The mean, standard deviation, median, min and max of the data are given below.

The explanatory variables are kilometres driven per year, zone, bonus, make of vehicle, and amount of time insured in policy years. The response variable is the claim frequency.

Figure 3.1: Density of Claims with a Gaussian Kernel and Bandwidth of 1.922.



The payment made is omitted for the claim frequency regressions and is considered later in the claim severity regression. Figure 3.1 depicts the density of claims with bandwidth calculated by what is known as Silverman's rule of thumb [Silverman, 2018]. The bandwidth is chosen in order to minimize the mean integrated squared error (MISE). MISE is given as $E = \int (f_n(x) - f(x))^2 dx$, where $f_n(x)$ is the estimate based on sample size n and $f(x)$ is the unknown density function.

4 Application of Claim Frequency, Severity and Expected Compensation

As described in Chapter 3, the compound Poisson, the compound Negative Binomial, the zero-inflated Poisson and the zero-inflated Negative Binomial distributions are used to model the claim frequencies. Claim severity follows a Gamma distribution. The claim frequency is modeled by the Poisson, Negative Binomial and the zero-inflated distributions and assumes independence from claim severity. In this Section, all the models are described. The primary result is expected compensation and is obtained by multiplying the expected frequency (from either of the Poisson or Negative Binomial models) by the expected claim severity (from the Gamma model). This is possible because expected compensation assumes independence between claim frequency and claim severity.

4.1 Claim Frequency

For claim frequency, the Poisson model, zero-inflated Poisson model, Negative Binomial model and zero-inflated Negative Binomial model are applied to the dataset. They are scored and compared using residual analysis and the DIC criterion. A fully Bayesian MCMC approach is used to find the posterior distribution of the parameters for both models respectively. Both the Poisson model and Negative Binomial model are tested with different groupings of covariates. The results are compared using the DIC criterion, and subsequently, the best model fits are retained.

4.1.1 Poisson and Zero-Inflated Poisson Models

The first models for claim frequency are the Poisson and Zero-inflated Poisson models. Considering only the observations with non-zero insured policy years, 1937 observations are obtained. The following model considers the number of claims $N_i, i = 1, 2, \ldots, n$, observed for each of the i^{th} insurance class categories:

$$N_i|\theta_i \sim Poisson(e_{it} * \theta_i^N), \tag{36}$$

where

$$\theta_i^N = exp(x_i'\beta).$$

Here, e_{it} denotes the insured policy years for policy holder i. The average number of claims is denoted by θ_i for class i, x_i is a vector of covariates, and β is a vector of parameters. To estimate the number of claims θ_i, the vector of unknown parameters is $\beta = (\beta_0, \beta_1, \ldots, \beta_4)'$. The vector of covariates for the i^{th} policy class observation is given by $x_i = (x_{0i}, x_{1i}, \ldots, x_{4i})$. The matrix obtained from above for the i^{th} observation is described in Table 4.1.

4.1.2 Negative Binomial and Zero-Inflated Negative Binomial Models

The second class of model chosen for claim frequency are the Negative Binomial and Zero-inflated Negative Binomial regression models. Just as with the Poisson model, the number of claims observed is $N_i, i = 1, 2, \ldots, n$ for J insurance class categories. From Equations (9) and (10),

$$N_i \sim NegativeBinomial(e_{it} * \theta_i^N, \sigma_i^2) \tag{37}$$

Table 4.1: Regression Covariates

Covariate (x)	Parameter (β)	Description
x_{0i}	β_0	Intercept
x_{1i}	β_1	Kilometres
x_{2i}	β_2	Zone
x_{3i}	β_3	Bonus
x_{4i}	β_4	Make

with mean θ_i^N given by

$$\theta_i^N = (exp(x_i'\beta)),$$

and variance given as

$$\sigma_i^2 = \theta_i(1 + \alpha\theta_i).$$

The matrix of regression covariates for the Negative Binomial model is the same as that for the Poisson regression covariates given in Table 4.1.

4.2 Claim Severity

4.2.1 Gamma Model

The proposed model for claim severity is the Gamma model. For this model, only observations with a positive number of claims is considered. With this considered, 1797

observations are obtained for the model. The covariance vector is the same as the model for claim frequency $x_i = (x_{i0}, x_{i1}, \ldots, x_{i4})'$. The vector of unknown regression parameters is $\eta = (\eta_0, \eta_1, \ldots, \eta_4)$. For policy holders $i = 1, 2, \ldots, n$, let $X_{ik}, k = 1, 2, \ldots, N_i$, denote the individual claim severity for the N_i observed claims. An individual claim severity conditional on N_i is assumed to be independently Gamma distributed:

$$X_{ik}|N_i \sim Gamma(\mu_i^S, v), k = 1, 2, \ldots, N_i, i = 1, 2, \ldots, n. \tag{38}$$

Since $X_{ik}|N_i, k = 1, 2, \ldots, N_i$ are assumed to be independent and identically distributed, the average claim severity X_i, given the observed number of claims N_i, is Gamma distributed with mean and variance

$$\begin{aligned}
&E(X_{ik}|N_i) = \frac{\mu_i^S}{v} \\
&Var(X_{ik}|N_i) = \frac{(\mu_i^S)^2}{v}, \\
&E\left[X_i = \sum_{k=1}^{N_i} \frac{X_{ik}}{N_i}|N_i\right] = \frac{\mu_i^S}{v} \\
&Var\left[X_i = \sum_{k=1}^{N_i} \frac{X_{ik}}{N_i}|N_i\right] = \frac{(\mu_i^S)^2}{vN_i}.
\end{aligned}$$

Therefore, the average claim severity for policy holder i is given by

$$X_i = \sum_{k=1}^{N_i} \frac{X_{ik}}{N_i}. \tag{39}$$

4.3 Expected Compensation

One of the primary goals of this thesis is to find the expected compensation cost insurance companies have to pay. As mentioned in Chapter 3, this cost depends on the claim

frequency and claim severity. The expected value of the total cost X_i per number of years insured e_i is a combination of either the Poisson, Zero-inflated Poisson, Negative Binomial or Zero-inflated Negative Binomial model with the Gamma model. The expected compensation can be expressed as

$$\begin{aligned}
E\left(\frac{X_{it}}{e_{it}}\right) &= \frac{1}{e_{it}}E(X_{it}) \\
&= \frac{1}{e_{it}}E\left(\sum_{k=1}^{N_{it}} W_{itk}\right) \\
&= \frac{1}{e_{it}}E(E(\sum_{k=1}^{N_{it}} W_{itk}|N_{it})) \\
&= \frac{1}{e_{it}}E(N_{it}E(W_{it1})) \\
&= \frac{1}{e_{it}}E(N_{it})E(W_{it1}) \\
&= \frac{1}{e_{it}}E(E(N_{it}|\theta_{it}^N))\mu_{it} \\
&= \frac{1}{e_{it}}E(e_{it}\theta_{it}^N)exp(x_i^T\eta) \\
&= exp(x_i^T\beta)exp(x_i^T\eta),
\end{aligned}$$

where $exp(x_i^T\beta)$ is the Poisson or Negative Binomial expected value and $exp(x_i^T\eta)$ is the Gamma expected value.

4.4 Residual Analysis

Evaluating model fit has been a controversial subject among Bayesian statistics. Bayesian prior to posterior inferences assume the whole structure of a probability model and can yield misleading inferences when the model is poor. A good Bayesian analysis, therefore,

should include at least some check of the adequacy of the fit of the model to the data and the plausibility of the model for the purposes for which the model will be used [Gelman et al., 2013]. According to [Gelman et al., 2013], it is necessary to examine models by their practical implications as well as tests for outliers, plots of residuals, and normal plots. In this thesis, both practical implications and residual plots are analysed. The practical checks implied are twofold. First, is the model consistent with the data? and second, do the inferences from the model make sense? Residual analysis, in particular graphical analysis, is an important statistical tool used to evaluate the quality of a model fit. Residuals are defined as the difference between the measured output from the data and the predicted model output. When errors in the residuals are not randomly distributed about zero, this suggests the model is not an appropriate fit.

A comprehensive residual analysis is employed. The residuals are plotted against the regressor variables, as well as the response variable to check for any outliers or curvature. Raw residuals are calculated by $r_i = y_i - \hat{\mu}_i$ and analysed. The pearson residual is also calculated to correct for the unequal variance in the residuals. The Pearson residual is given by

$$p_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\phi}\hat{\mu}_i}},$$

where

$$\hat{\phi} = \frac{1}{n-k}\sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}.$$

For the Negative Binomial regression, the Pearson residual is given as

$$p_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i + \alpha \hat{\mu}_i{}^2}}.$$

Finally, the hat values are calculated. The hat matrix is used to measure the influence of each observation. The hat values, h_{ii}, are the diagonal entries of the hat matrix calculated using

$$H = W^{1/2} X (X'WX)^{-1} X' W^{1/2},$$

where W is a diagonal matrix with $\hat{\mu}_i$. That is, a square matrix with $\hat{\mu}_i$ along the diagonal, and zeroes elsewhere. Residual analysis for the claim severity (with Gamma regression model) has a standardized ordinal residual defined by

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{Var(y_i)}}}, \tag{40}$$

where

$$\widehat{Var(y_i)} = \frac{\hat{\mu}_i{}^2}{\hat{\alpha}_i}.$$

4.5 Bayesian Inference

4.5.1 Prior Distributions

An important aspect of Bayesian sampling and MCMC inference is the judicious selection of prior distributions. It can be an asset to have prior belief of the probability distributions of the parameters before the data is examined. With this knowledge, a subjective prior can be chosen. Subjective priors have a significant impact on the posterior distribution.

On the other hand, objective priors are selected to have a minimal impact on the posterior distribution. It can be said that objective priors are used to obtain results which are objectively valid. This means the results completely depend on the data. The priors selected for these models are flat priors. Flat priors place more weight on the likelihood function, thus have limited impact on the posterior distributions.

Poisson Models

$$\beta[i] \sim N(\mu_\beta, \sigma_\beta^2) \qquad \mu_\beta = 0, \qquad \sigma_\beta^2 = 1000$$

Negative Binomial Models

$$\begin{aligned} &\beta[i] \sim N(\mu_\beta, \sigma_\beta^2) && \mu_\beta = 0, \quad \sigma_\beta^2 = 1000 \\ &\alpha \sim U(a, b) && a = 0, \quad b = 50 \end{aligned}$$

Gamma Model

$$\begin{aligned} &\eta[i] \sim N(\mu_\beta, \sigma_\beta^2) && \mu_\beta = 0, \quad \sigma_\beta^2 = 1000 \\ &\alpha \sim U(a, b) && a = 0 \quad b = 100 \end{aligned}$$

In the foregoing, N represented a normal distribution and U represented a uniform distribution.

4.5.2 Posterior Distribution

The posterior distribution of the parameters, given the observed claim numbers and claim severity, describe the uncertainty and is the result used for inference in Bayesian analysis.

The basis of Bayesian statistics is Bayes' theorem, as was seen in Equation (30). In this case, the prior is uninformative and the posterior results will be data driven for all three models. The posterior for the Poisson distribution is

$$\begin{aligned} p(\theta_i^N, \beta_0, \ldots, \beta_4 | N_i) &= \frac{p(\theta_i^N, \beta_0, \ldots, \beta_4 | N_i)}{p(N_i)} \\ &\propto p(\theta_i^N, \beta_0, \ldots, \beta_4 | N_i) = p(N_i | \theta_i^N) p(\theta_i^N | \beta_0, \ldots, \beta_4) p(\beta_0) \ldots p(\beta_4) \end{aligned}$$

where $f(N_i)$ is the prior and $p(N_i|\theta_i^N)$ is the likelihood. The flat prior is considered to be insignificant, since it expresses vague information about the variable. Since the likelihood function yields more information than the uninformative prior, the posterior is proportional to the likelihood of the distribution $p(N_i|\theta_i^N)$. The posterior to the Negative Binomial follows the same rules as the Poisson posteriors, and is given as

$$\begin{aligned} p(\theta_i^N, \beta_0, \ldots, \beta_4, \alpha | N_i) &\propto p(N_i | \theta_i^N) p(\theta_i^N | \beta_0, \ldots, \beta_4, \alpha) p(\beta_0) \ldots p(\beta_4) \\ p(\theta_i^N, \beta_0, \ldots, \beta_4, \alpha | N_i) &= \frac{p(\theta_i^N, \beta_0, \ldots, \beta_4, \alpha | N_i)}{p(N_i)} \\ &\propto p(\theta_i^N, \beta_0, \ldots, \beta_4, \alpha | N_i) \\ &= p(N_i | \theta_i^N) p(\theta_i^N | \beta_0, \ldots, \beta_4, \alpha) p(\beta_0) \ldots p(\beta_4) p(\alpha) \\ p(\theta_i^N, \beta_0, \ldots, \beta_4, \alpha | N_i) &\propto p(N_i | \theta_i^N) p(\theta_i^N | \beta_0, \ldots, \beta_4, \alpha) p(\beta_0) \ldots p(\beta_4) p(\alpha) \end{aligned}$$

Lastly, the posterior distribution for claim severity is

$$\begin{aligned} p(\eta_0, \ldots, \eta_4, \alpha | X_{ik}) &= \frac{p(\eta_0, \ldots, \eta_4, \alpha | X_{ik})}{p(X_{ik})} \\ &\propto p(\eta_0, \ldots, \eta_4, \alpha | X_{ik}) \\ &= p(X_{ik} | N_{ik}) p(N_{ik} | \eta_0, \ldots, \eta_4, \alpha) p(\eta_0) \ldots p(\eta_4) p(\alpha) \end{aligned}$$

where $f(X_{ik})$ is the prior and $p(X_{ik}|N_{ik})$ is the likelihood for claim severity. Once more, the prior is considered agnostic and therefore the posterior distribution is proportional to the likelihood of the distribution $p(X_{ik}|N_{ik})$.

4.6 Markov Chain Monte Carlo Application

As mentioned in Section 2.7.1, the Gibbs sampler is used as a form of MCMC simulation. The Gibbs sampler provides samples from the posterior distributions. Three chains are run simultaneously in order to compute variances. Chains are run for varying simulations of length n in order to enhance the chances of convergence. By sampling from the posterior distribution of the Gibbs sampler, the conditional distribution for all parameters are obtained. The Gibbs MCMC estimates the mean, standard deviation, naïve standard error, and time-series standard error. These results are given in Section 5.

4.6.1 MCMC Calculations

The program used in R which runs Gibbs sampling produce many computations. These computations are explained in this Section. The mean from the Results is the average expected value of the unknown parameters β across all iterations of the program. The standard deviation calculated is the standard deviation of the mean. This is the variance or dispersion of the mean of each iteration. The naïve standard error is the standard error of the mean adjusted for sample size which captures the simulation error of the mean rather than the posterior uncertainty. The time-series standard error adjusts the naïve

standard error for autocorrelation. This is to say that the time-series standard error is calculated by taking the average of the variance of each set of samples of each chain.

The Credible Intervals for each model are also calculated. Credible intervals are the Bayesian equivalent of the confidence interval. Confidence intervals express uncertainty in the knowledge with a range designed to include the true parameter with some probability, commonly 95%. This confidence interval is interpreted in the way that if 100 experiments are run, 95 of them will be at least within the interval width. A Bayesian could criticize the confidence interval since the 5% of the results not in the confidence interval can be nonsense, just as long as the 95% are within the confidence interval. Furthermore, a Bayesian could say that the only experiment that matters is the one being ran, not the other 99 to test the confidence interval. Bayesian's approach a parameter as fixed with some probability distribution. The credible interval can be interpreted as an 95% chance of having a parameter within the interval. For example, if the credible interval of 95% for the average height of students at Laurentian is between 160 and 180 centimeters, this means there is a 95% chance the average height of a student is within the interval.

The claim frequencies and severities are primarily calculated by area. This is to say, the results given is the average claim frequency/severity for each of the stated areas.

Autocorrelation is commonly defined as the degree of similarity between a given time-series and a lagged version of itself. Since the $(i+1)^{th}$ iteration depends only on the i^{th} iteration, autocorrelation is often quite high for Gibbs sampling. Autocorrelation with a lag of 50 is the degree of similarity between the $(i+50)^{th}$ iteration and the i^{th} simulation.

The high autocorrelation goes against the Gibbs sampler used. It is a strong indicator that running longer chains is required. Rather than thinning the chains and throwing away samples, the chains are ran for longer when autocorrelation is high as described in Section 2.7.1.

5 Results

In this chapter the results from all the methods employed are shown. To start, we model the expected claim frequency per year insured. Second, the expected claim severity per year insured. Finally, the expected compensation cost per year insured. The regression for each of the models are also given. Each model is analysed with residual analysis and scored with the DIC criterion. Trace plots and analysis for the MCMC chains are given.

5.1 Claim Frequency

5.1.1 Poisson Model

The Poisson model obtained after running 30,000 iterations of the Poisson is given by

$$E\left(\frac{\theta_i}{t_i}\right) = exp(-1.89657 + 0.14532x_1 - 0.10590x_2 - 0.19687x_3 - 0.03691x_4) \qquad (41)$$

Since this model, as with the other models given in this Section, is obtained by Bayesian inference, the posterior distribution can be directly examined to see which parameter values are most credible. Unlike in frequentist statistical analysis, there is no need to generate sampling distributions from null hypotheses and to figure out the probability that

fictitious data would be more extreme than the observed data. In other words, there is no need for p values and p value based confidence intervals [Kruschke and Liddell, 2018]. Instead, measures of uncertainty are based directly on posterior credible intervals provided in Table 5.3. The trace plots for the MCMC Gibbs samplers are given in Figure 5.1. We see that the criterion given by [Toft et al., 2007] listed in Section 2.7 are satisfied. The plots show that each variable mixed well, in that, it exhibits a rapid up and down variation with no long-term trends or drifts. The chain has mild correlation between draws and explores the sample space many times. Figure 5.1 gives both traceplots and density plots for the Poisson iterations. As true for all traceplots in this Section, traceplots represent each sample step from the iterations ran. As each iterations jumps from one value to the next around the mean of the parameter, the traceplot allows for analysis of how well the parameter traveled in the state space. The traceplots do not show any significant changes in the target distributions. The density plots given in Figure 5.1 are the smoothed histograms of their respectful traceplots. That is to say, they represent the posterior distribution of the parameter. The normal curve is important for an accurate distribution about the mean.

HISTOGRAM

Furthermore, the autocorrelation plots are given in Figure 5.2. These plots support the likelihood of convergence from Figure 5.1. The relatively low autocorrelation after lag $<$50 suggests fast convergence for the model.

Figure 5.1: Trace Plots and Posterior Density Plots for Poisson Iterations. 3 Chains With 30,000 Iterations Each

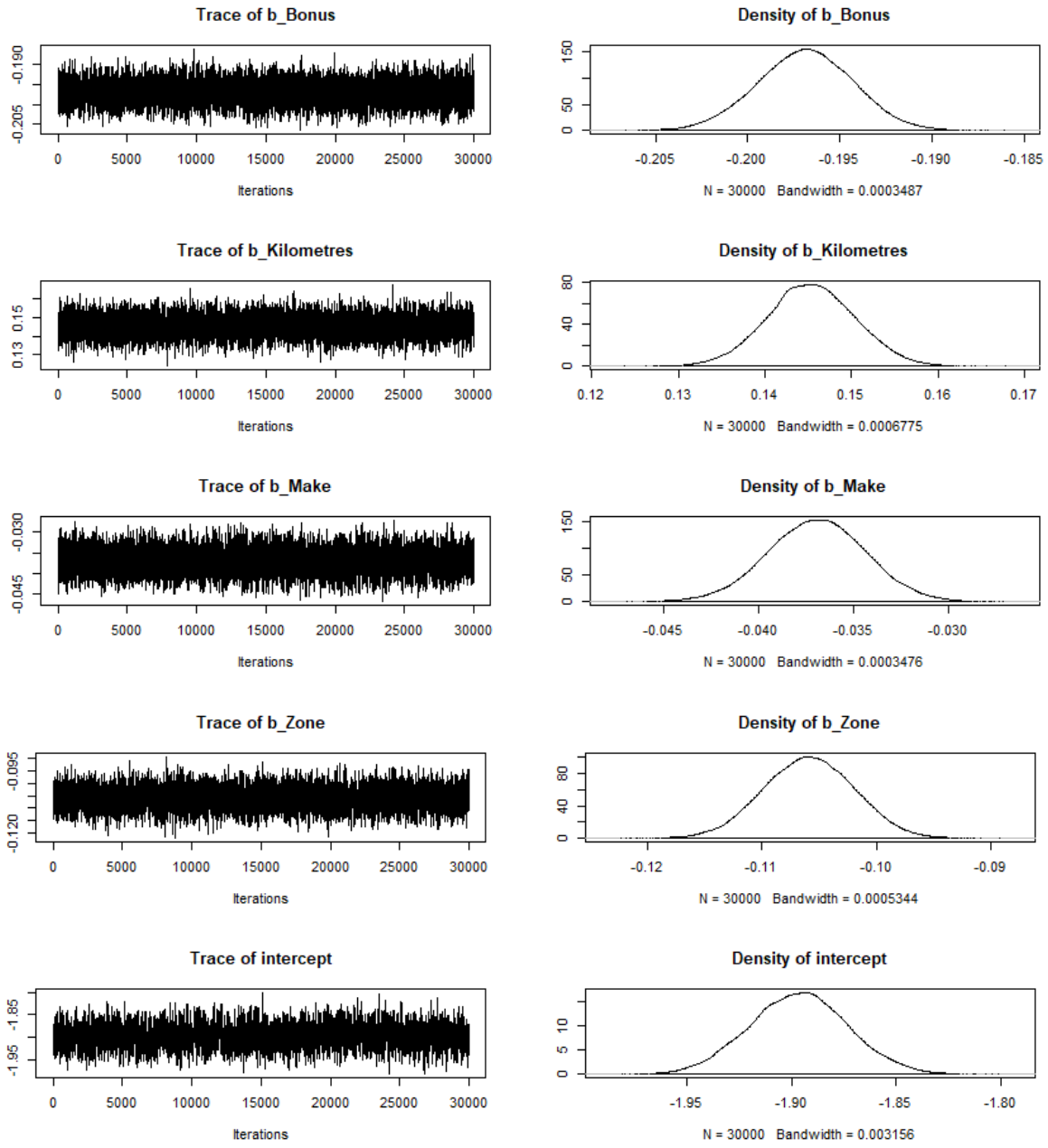


Figure 5.2: Autocorrelation Plot for Poisson Iterations

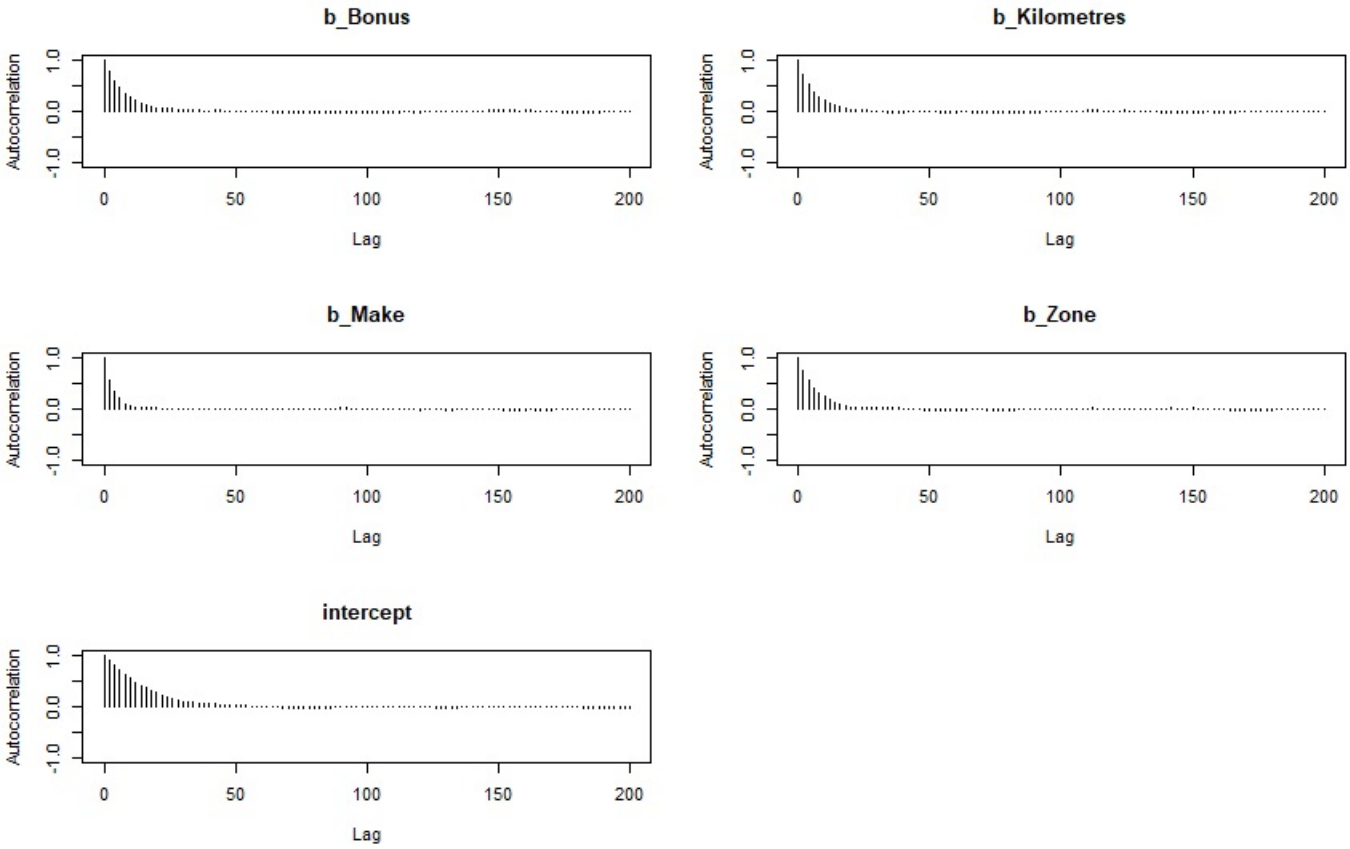


Table 5.1: Poisson Heidelberg and Welch diagnostics

	Stationarity and Half-width Test	Start iteration	P-value	Mean	Half-width
Intercept	Passed	1	0.0772	-1.896	1.08e-03
Bonus	Passed	1	0.7958	-0.197	8.05e-05
Kilometres	Passed	24001	0.0574	0.145	1.91e-04
Make	Passed	1	0.2974	-0.037	5.90e-05
Zone	Passed	1	0.2728	-0.106	1.22e-04

Table 5.1 gives the Heidelberg and Welch diagnostics. All the variables passed the stationarity and half-width test. Aside from kilometres, the variables all passed the stationarity test throughout the chain. The kilometres variables discarded the first 24,000 iterations of the chain before passing the stationarity test as described in Section 2.7.1. The large P-values obtained in the Heidelberg and Welch diagnostics are a good indication that the null hypothesis cannot be rejected and the chains are from a stationary distribution as described in Section 2.7.2. These p-values are not significant levels for the variables in the Poisson regression, but significance levels for the null hypothesis, the Gibbs sampling chains are from a stationary distribution.

The results of the full MCMC iterations for all three chains is given in Table 5.2.

Table 5.2 gives the mean, standard deviation, naïve SE and time-series SE of the pos-

Table 5.2: Poisson Posterior Results

	Mean	SD	Naïve SE	Time-series SE
Intercept	-1.89657	0.023401	1.351e-04	5.390e-04
Bonus	-0.19687	0.002585	1.493e-05	4.113e-05
Kilometres	0.14532	0.005048	2.914e-05	7.332e-05
Make	-0.03691	0.002578	1.488e-05	3.168e-05
Zone	-0.10590	0.003963	2.288e-05	6.107e-05

terior distribution calculated using the Poisson regression by means of Gibbs sampling. The posterior quantiles for each variable are given in Table 5.3. These are the confidence intervals for each estimated variable. From these results, it is seen that the Bonus and Kilometres have the most significant impact on the expected frequency. This is because the means for both of these variables are larger than their counterparts. Since the calculations in Table 5.2 are calculated by taking the average across all iterations ran, the small standard deviation across all variables indicates a small dispersion of the data for each estimate of the model. The small standard error across the board indicates that the mean is a good reflection of the actual mean. The credible intervals for the lower and upper bounds of the parameter estimates are given in Table 5.3

The closeness of the interval at 95% is ideal and signifies a fairly good fit for the model.

Table 5.3: Poisson Credible Intervals

	2.5%	25%	50%	75%	97.5%
Intercept	-1.94236	-1.91237	-1.8965	-1.88086	-1.85007
Bonus	-0.20198	-0.19859	-0.1968	-0.19510	-0.19185
Kilometres	0.13538	0.14199	0.1453	0.14872	0.15527
Make	-0.04205	-0.03865	-0.0369	-0.03518	-0.03189
Zone	-0.11367	-0.10859	-0.1059	-0.10320	-0.09823

The 95% interval can be interpreted as the values between the 2.5% and 97.5% values. For example, the 95% confidence interval for the Bonus coefficient is -0.20198 to -0.19185. Moving forward with the results, we can now interpret the expected claim frequencies per area. The results for the aforementioned are given in Table 5.4.

The inferences from the Poisson model, along with those of the other models as seen in the rest of Section 5, are in accordance with former knowledge of Claim frequencies. The results obtained from Table 5.4 are in keeping with what is expected. The higher claim frequencies come from larger cities with busier streets and more traffic. In the Appendix, Table 7.1 gives credible intervals for the claim frequency by zone. The average 95% error interval for the claim frequency by zone is 0.0191. The lower bound of the estimates vary by 13.73% from the mean, while the upper bound vary by 15.93% from the mean. For fur-

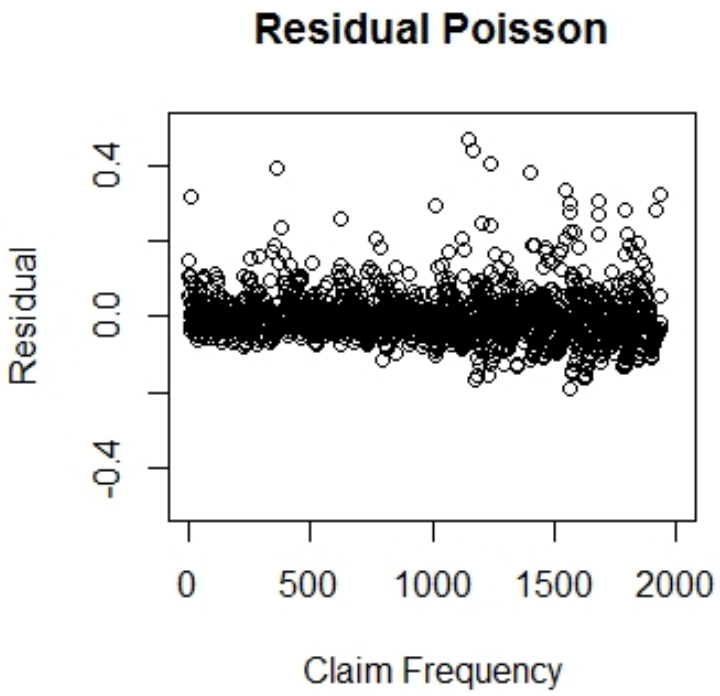
Table 5.4: Claim Frequency by Area (Poisson Results)

Code	Zone	Claim Frequency
1	Stockholm, Goteborg, Malmo with surroundings	0.08899375
2	Other large cities and surroundings	0.08005118
3	Small cities in northern Sweden	0.0720072
4	Small cities in southern Sweden	0.06477153
5	Rural areas in northern Sweden	0.05821204
6	Rural areas in southern Sweden	0.05240836
7	Gotland	0.04524769

ther results, see Table 7.2 for the claim frequency by zone and kilometres driven. As the second practical check, the model is consistent with the data. Replicated data generated under the model does look similar to the previously observed data.

The residual plot for the Poisson model is analysed and given in Figure 5.3. The random pattern about zero represents the stochastic error of the model and indicates an appropriate fit from the Poisson model. The slightly positive trend in some of the residuals could be an indication of the strong influence the larger claim frequencies have on the model. The Poisson model is said to be over-dispersed, since the residual deviance/degrees of freedom is > 1. As is often the case with rare event count data, this suggests that the Poisson model is not the optimal model. Therefore, as previously mentioned in Section 5.1.3, the Negative Binomial model is analysed.

Figure 5.3: Residual Plot for Poisson Model



5.1.2 Zero-Inflated Poisson Model

The Zero-inflated Poisson model required many more iterations to achieve convergence. After running 300,00 iterations on 3 chains simultaneously, the model obtained is given by

$$E\left(\frac{\theta_i}{t_i}\right) = exp(-1.89501 + 0.14525x_1 - 0.10605x_2 - 0.19699x_3 - 0.03697x_4). \quad (42)$$

Equation (42) is the Zero-inflated regression model for the expected claim frequency per year of coverage. The slow convergence of the Markov chain can be caused by bad starting values, high posterior correlation, or under-parameterized models to name a few reasons. It is concluded that the slow convergence is a result of assuming no prior knowledge and the limited parameters for the model. In further work, the issue can be addressed with more data and prior knowledge to select informative priors.

As we can see from the Trace plots, all of the Zero-inflated Poisson variables mixed rather well aside from the intercept (β_0). For this reason, the chains were run much longer to ensure convergence.

The higher autocorrelation of β_0 was expected since it was slower to converge. However, the chains still tend towards convergence as seen from the Heidelberg and Welch diagnostic in Table 5.5. All variables passed both test and no iterations were discarded from the first half of the test. The large p-value from the intercept is a good indication that the null hypothesis cannot be rejected and the chains are from a stationary distribution.

Figure 5.4: Trace Plots and Posterior Density Plots for Zero-Inflated Poisson Iterations. 3 Chains With 300,000 Iterations Each

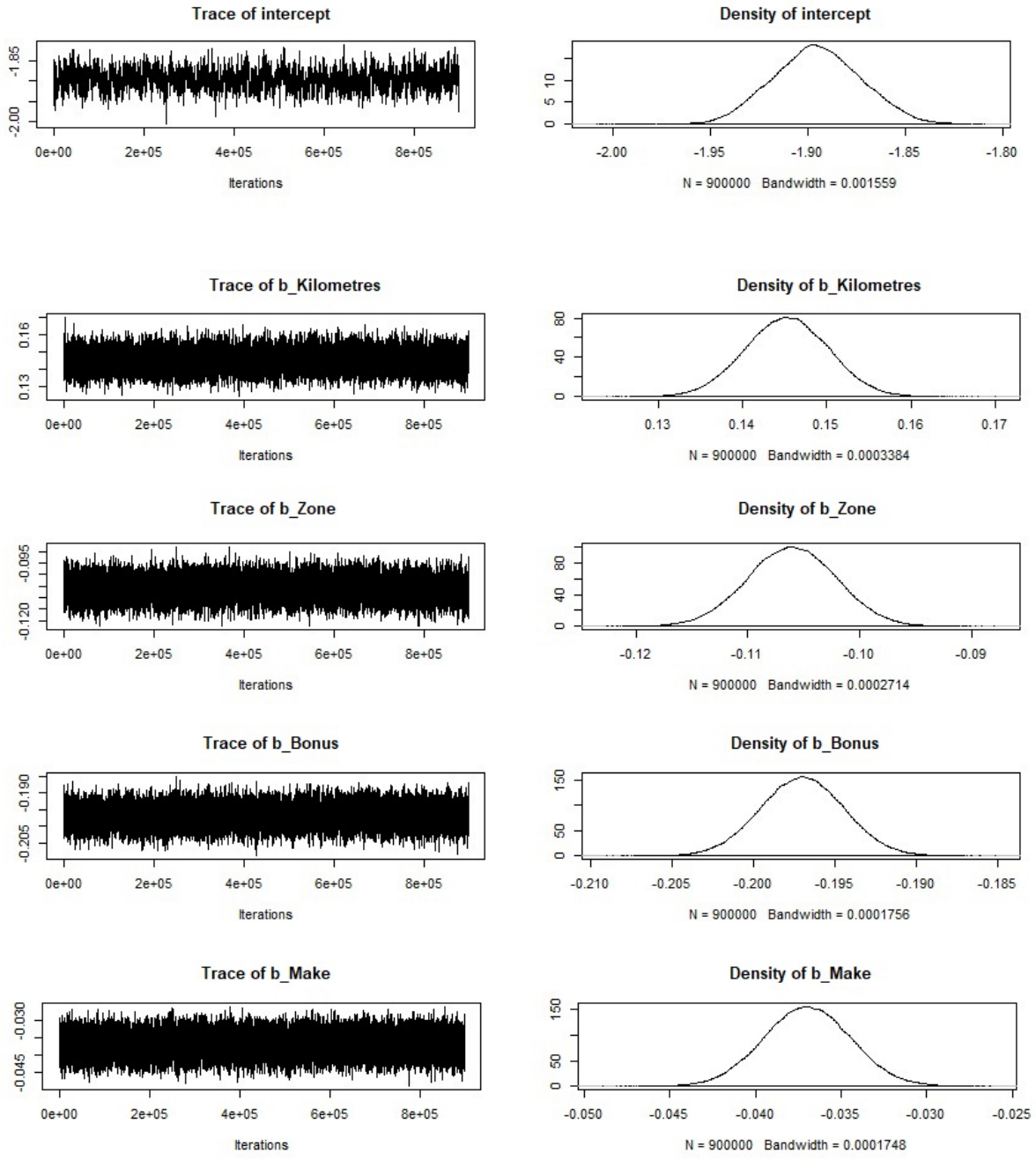


Figure 5.5: Autocorrelation Plot for Zero-Inflated Poisson Iterations

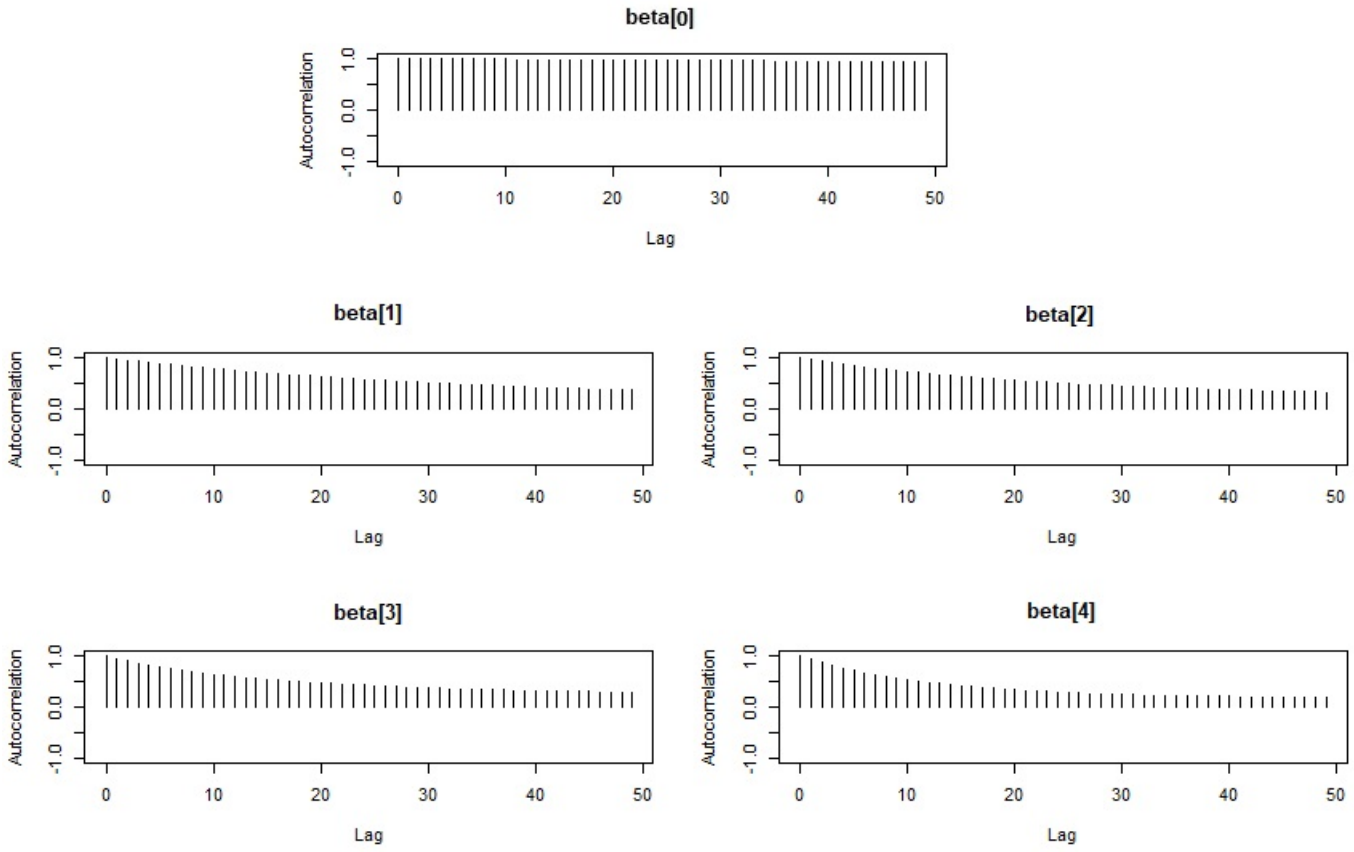


Table 5.5: Zero-Inflated Poisson Heidelberg and Welch diagnostics

	Stationarity and Half-width Test	Start iteration	P-value	Mean	Half-width
Intercept	Passed	1	0.974	-1.8962	3.17e-03
Bonus	Passed	1	0.299	-0.1969	8.52e-05
Kilometres	Passed	1	0.102	0.1454	1.81e-04
Make	Passed	1	0.257	-0.0369	6.78e-05
Zone	Passed	1	0.271	-0.1060	1.44e-04

Table 5.6: Zero-Inflated Poisson Posterior Results

	Mean	SD	Naïve SE	Time-Series SE
Intercept	-1.89501	0.022828	2.406e-05	9.133e-04
Bonus	-0.19699	0.002571	2.710e-06	2.500e-05
Kilometres	0.14525	0.004954	5.222e-06	5.340e-05
Make	-0.03697	0.002559	2.697e-06	2.003e-05
Zone	-0.10605	0.003973	4.187e-06	4.133e-05

Table 5.7: Zero-Inflated Poisson Credible Intervals

	2.5%	25%	50%	75%	97.5%
Intercept	-1.93856	-1.9105	-1.89526	-1.87967	-1.85032
Bonus	-0.20204	-0.1987	-0.19698	-0.19525	-0.19195
Kilometres	0.13556	0.1419	0.14526	0.14861	0.15497
Make	-0.04196	-0.0387	-0.03698	-0.03525	-0.03194
Zone	-0.11386	-0.1087	-0.10605	-0.10335	-0.09830
r (shape)	9.65539	10.61896	11.16488	11.73455	12.91373

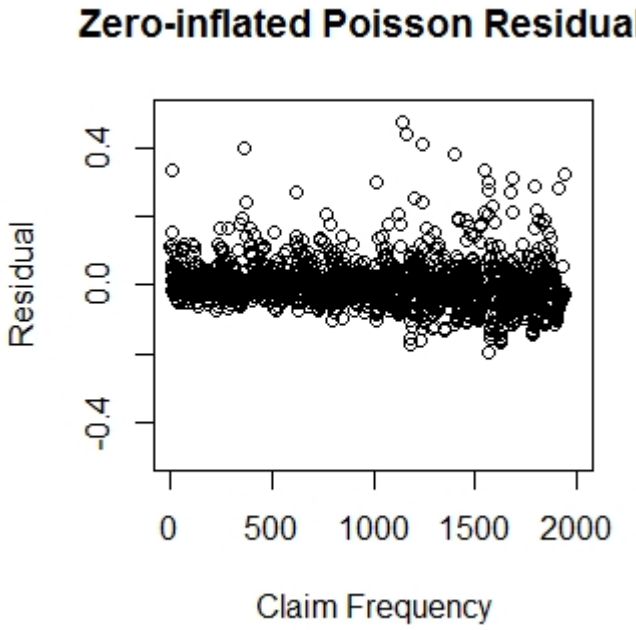
The credible intervals are narrow with relatively small standard deviations as seen in Table 5.6. This signifies that the true estimation of the mean is relatively precise. See Table 5.8 for the results of the claim frequency from the Zero-inflated Poisson model.

The claim frequency results from the Zero-inflated model are comparable to the results obtained from the Poisson model. This indicates that the results are an appropriate fit since both models produced similar results. The data does not have enough zero-valued claims to have a noticeable improvement in the model and results. Therefore, a Zero-inflated model is not necessary since it has a much higher computing time to obtain the same results. Nevertheless, it is good to have a second model to complement the Poisson results. The 95% credible interval for the Zero-inflated Poisson model has an average

Table 5.8: Claim frequency by zone (Zero-Inflated Poisson)

Code	Zone	Claim Frequency
1	Stockholm, Goteborg, Malmo with surroundings	0.0890412
2	Other large cities and surroundings	0.08008185
3	Small cities in northern Sweden	0.07202399
4	Small cities in southern Sweden	0.06477691
5	Rural areas in northern Sweden	0.05820826
6	Rural areas in southern Sweden	0.05239699
7	Gotland	0.04523124

Figure 5.6: Residual Plot for Zero-Inflated Poisson Model



upper-bound that is 9.26% away from the mean. The average lower-bound of the same credible interval is 7.77% away from the mean. In the Appendix, Table 7.3 gives the full credible intervals for the claim frequency by zone for the Zero-inflated Poisson model. The random pattern about zero in the residual plot from Figure 5.6 signifies an appropriate model fit.

5.1.3 Negative Binomial Model

The Negative Binomial model obtained after running 50,000 iterations is given by

$$E\left(\frac{\theta_i}{t_i}\right) = exp(-1.86302 + 0.11447x_1 - 0.10066x_2 - 0.19096x_3 - 0.03691x_4). \qquad (43)$$

Equation (43) is the expected claim frequency per year of coverage. These values represent the posterior mean of each variable x_i.

The trace plots from the Negative Binomial MCMC iterations are analysed. The Negative Binomial converged significantly slower than its Poisson counterpart. For this reason, the chains ran longer. The autocorrelation plot for the Negative Binomial is also considered for convergence, see Figure 5.8.

We see much greater autocorrelation in these plots which is not conclusive for precision of estimates. However, with a long enough chain, the precision is improved and autocorrelation becomes insignificant. In Table 5.9, the passed stationarity and half-width test results are shown. None of the iterations were discarded for any of the variables since the null hypothesis could not be rejected from the first iteration.

The credible intervals for the Negative Binomial posterior are bigger than the Poisson posteriors which is to be expected. The mean of the Negative Binomial is slow to converge towards a normal distribution [Shilane et al.,] . This suggests that the Negative Binomial chain requires a larger sample size to produce more accurate results. To obtain more conclusive predictive estimators, the chain would ideally be run again with a larger data set.

The results obtained for Claim Frequency from the Negative Binomial posteriors are slightly different than those from the Poisson posteriors. However, they still follow the same trend of higher claim frequencies in larger cities. The Negative Binomial claim frequencies are higher than those from the Poisson model. In the Appendix, Table 7.5 gives

Figure 5.7: Trace Plots and Posterior Density Plots for Negative Binomial Iterations.

3 Chains With 75,000 Iterations Each

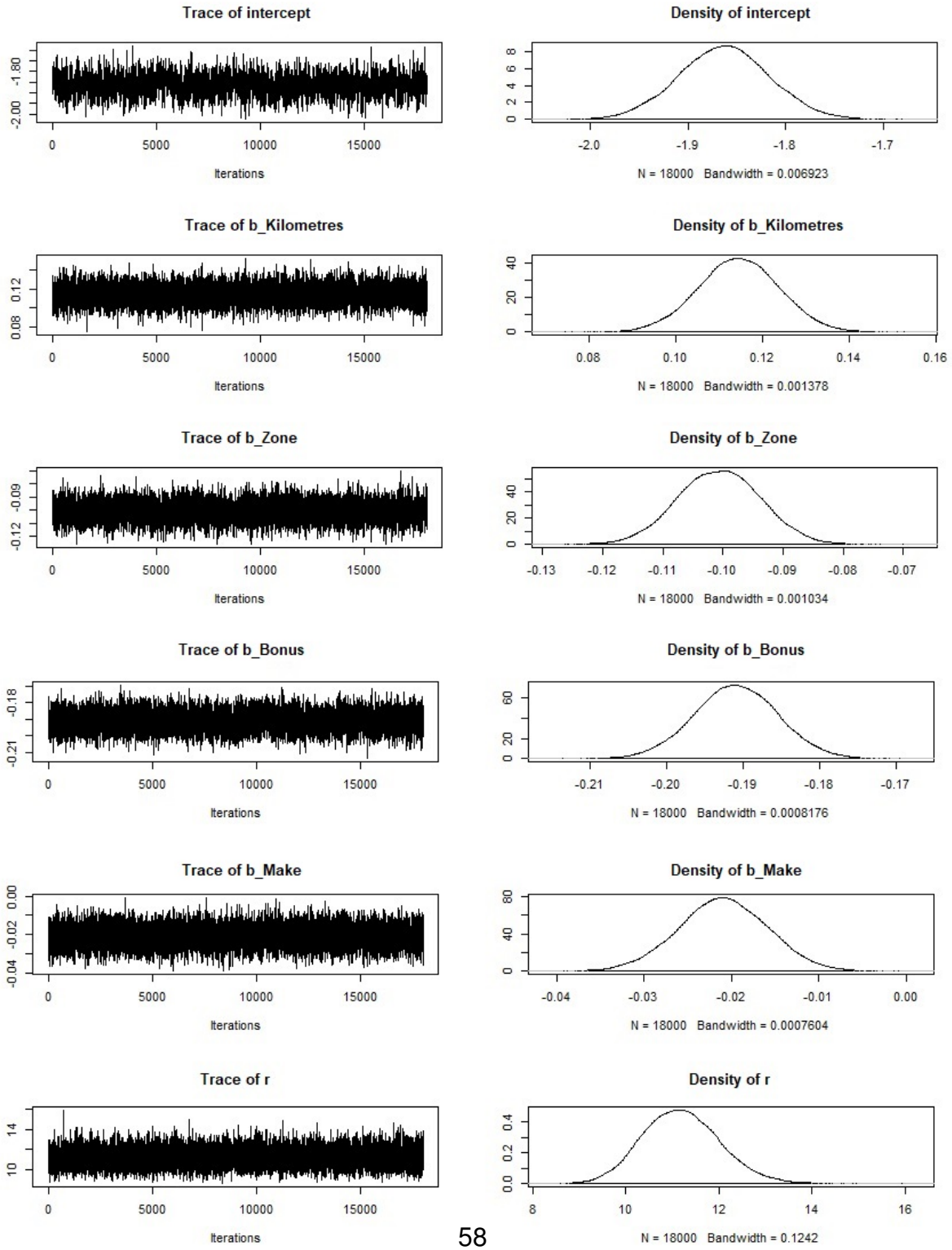
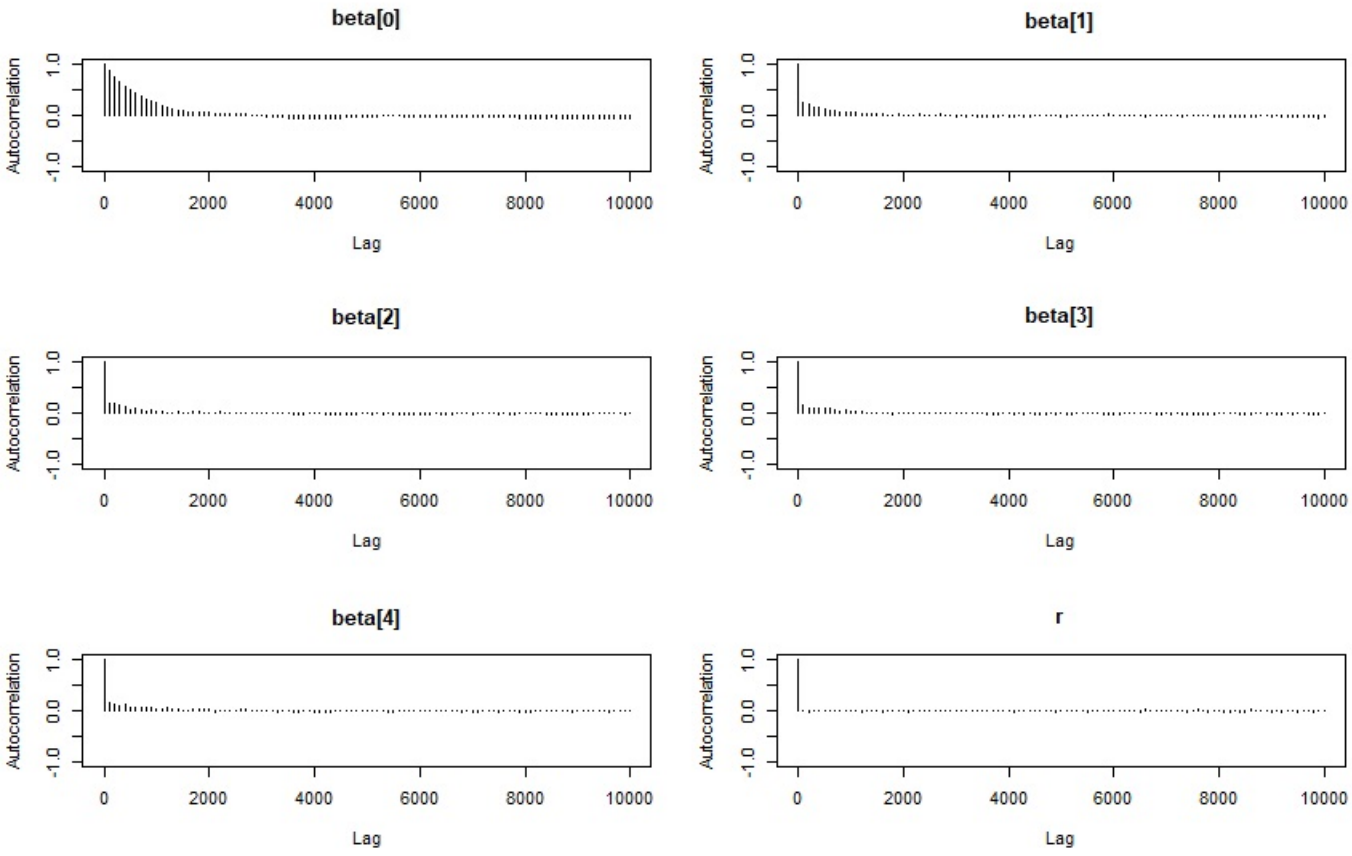


Figure 5.8: Autocorrelation Plot for Negative Binomial Iterations



credible intervals for the claim frequency by zone. The average 95% error interval for the claim frequency by zone is 0.0386. The lower bounds vary on average by 24.56% from the mean, while the upper bounds vary by 32.74% from the mean. The large credible interval is of concern for the accuracy of the estimates. A larger sample size would be needed for more conclusive results. For further results on claim frequencies, see Table 7.2 for the claim frequency by zone and kilometres driven.

The residual plot for the Negative Binomial is given in Figure 5.9. The random pattern about zero for the residuals of the Negative Binomial model is supportive of the model fit and thus is not a cause for concern.

Figure 5.9: Residual Plot for Negative Binomial Model

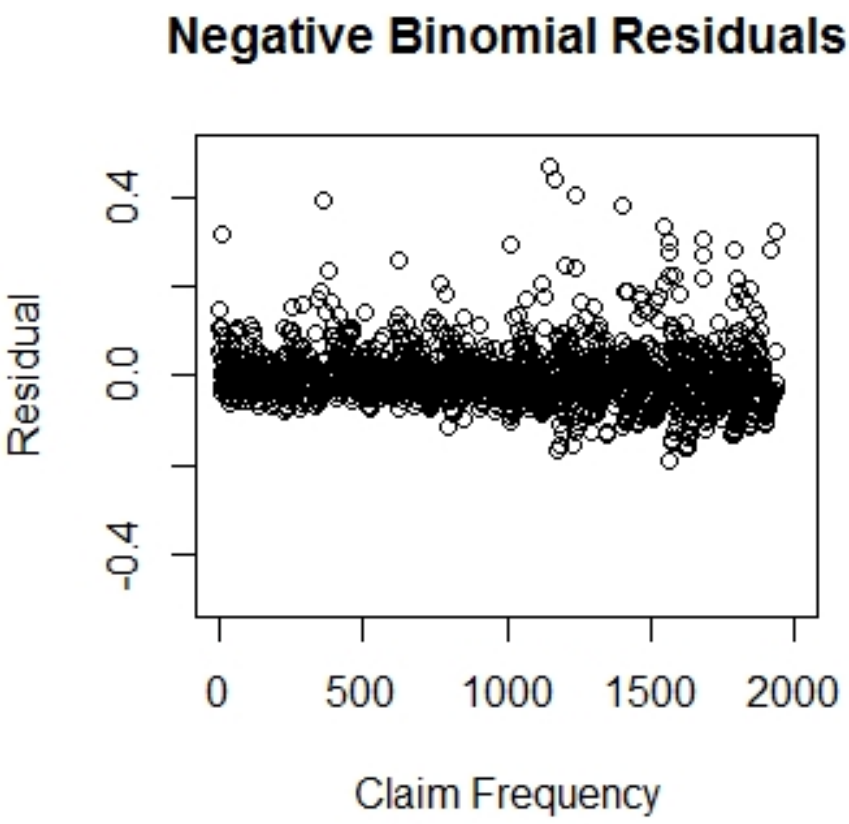


Table 5.9: Negative Binomial Heidelberg and Welch diagnostics

	Stationarity and Half-width Test	Start iteration	P-value	Mean	Half-width
Intercept	Passed	1	0.883	-1.8641	3.38e-03
Bonus	Passed	1	0.209	-0.1910	9.16e-05
Kilometres	Passed	1	0.362	0.145	1.91e-04
Make	Passed	1	0.2974	-0.0209	8.00e-05
Zone	Passed	1	0.155	-0.1006	1.32e-04
r	Passed	1	0.198	11.2002	2.72e-03

Table 5.10: Negative Binomial Posterior Results

	Mean	SD	Naïve SE	Time-Series SE
Intercept	-1.86302	0.046429	3.46E-04	1.37E-03
Bonus	-0.19096	0.005474	4.08E-05	7.38E-05
Kilometres	0.11447	0.009265	6.91E-05	1.52E-04
Make	-0.02096	0.005091	3.79E-05	6.95E-05
Zone	-0.10066	0.006922	5.159e-05	1.07E-04
r (shape)	11.19543	0.831351	6.20E-03	6.20E-03

Table 5.11: Negative Binomial Credible Intervals

	2.5%	25%	50%	75%	97.5%
Intercept	-1.95364	-1.89419	-1.86309	-1.83208	-1.77118
Bonus	-0.20175	-0.19465	-0.19096	-0.18726	-0.18018
Kilometres	0.09614	0.10832	0.11448	0.12069	0.13252
Make	-0.03109	-0.02433	-0.02097	-0.01750	-0.01103
Zone	-0.11421	-0.10534	-0.10063	-0.09599	-0.08723
r (shape)	9.65539	10.61896	11.16488	11.73455	12.91373

Table 5.12: Claim Frequency by Zone (Negative Binomial Results)

Code	Zone	Claim Frequency
1	Stockholm, Goteborg, Malmo with surroundings	0.09141397
2	Other large cities and surroundings	0.08266021
3	Small cities in northern Sweden	0.0747447
4	Small cities in southern Sweden	0.06758718
5	Rural areas in northern Sweden	0.06108674
6	Rural areas in southern Sweden	0.05526271
7	Gotland	0.0480894

5.1.4 Zero-Inflated Negative Binomial Model

The Zero-inflated Negative Binomial model obtained after running 3 chains for 400,000 iterations each is given by

$$E\left(\frac{\theta_i}{t_i}\right) = exp(-1.86180 + 0.11435x_1 - 0.10075x_2 - 0.19105x_3 - 0.02099x_4). \quad (44)$$

Equation (44) is the regression for the claim frequency obtained from the Zero-inflated Negative Binomial model. The Zero-inflated Negative Binomial converged significantly slower than the other models tested. The intercept, β_0 did not mix as well as the other variables. The autocorrelation plots and the Heidelberg and Welch test results are given in Table 5.11 and Table 5.13 respectively.

The high autocorrelation means the chains will need to be run longer than usual. This is why iterations of length 400,000 are chosen. The Heidelberg and Welch tests passed for all the variables in the model. The first 10% of bonus, kilometres, make and zone estimators are discarded since the null hypothesis that the chain is from a stationary distribution is rejected for the entire chain. For the rest of the chain (90%), the null hypothesis cannot be rejected and so this chain is from a stationary distribution.

The credible intervals for the Zero-Inflated Negative Binomial posteriors are similar to the Negative Binomial. Once again, if possible the Zero-inflated Negative Binomial chain should be run with a larger sample size to produce more conclusive results. The results from Table 5.16 strongly support those from the Negative Binomial in Table 5.12. This also

Figure 5.10: Trace Plots and Posterior Density Plots for Zero-Inflated Negative Binomial Iterations. 3 Chains With 400,000 Iterations Each

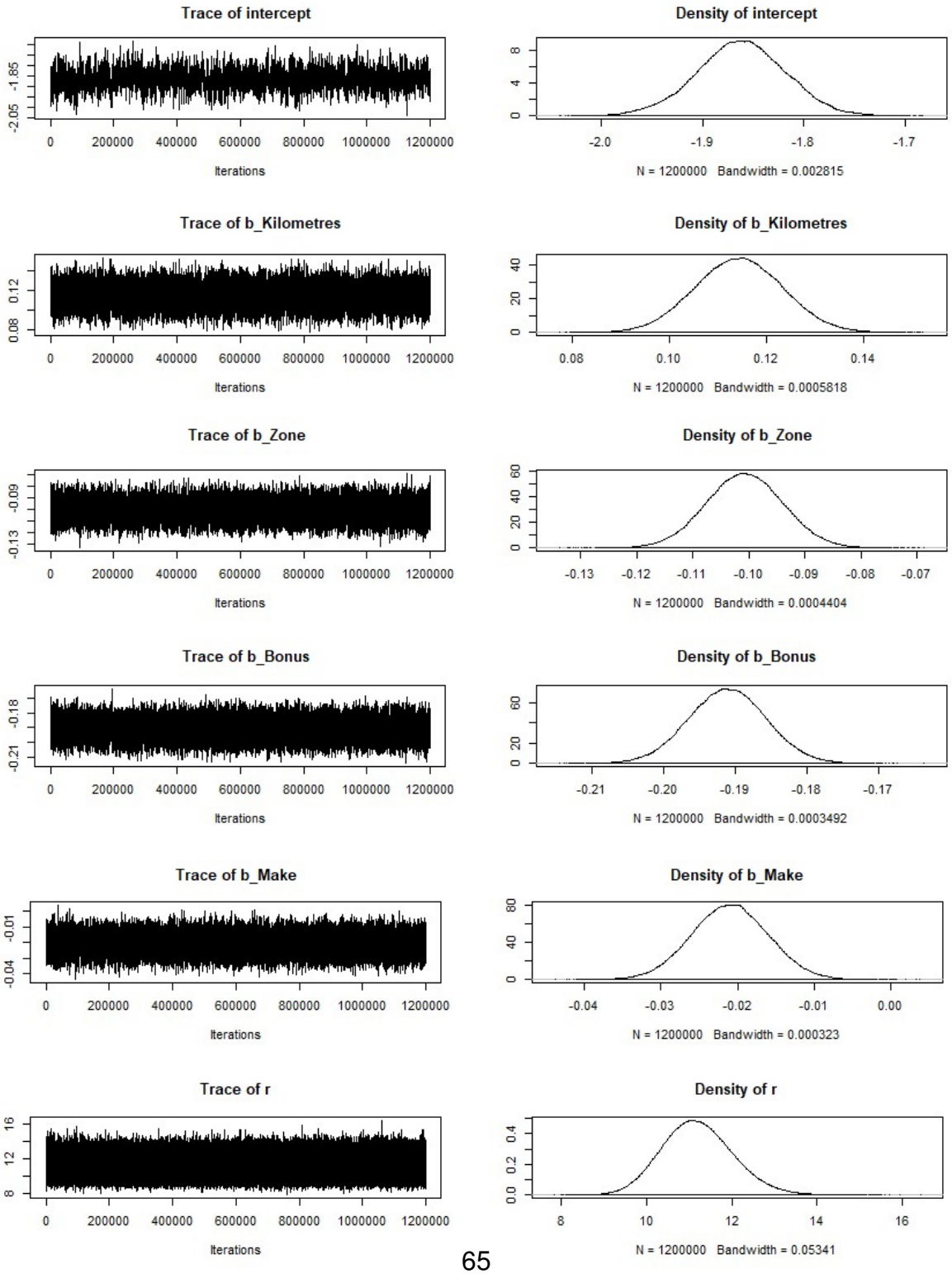


Figure 5.11: Autocorrelation Plot for Zero-Inflated Negative Binomial Iterations

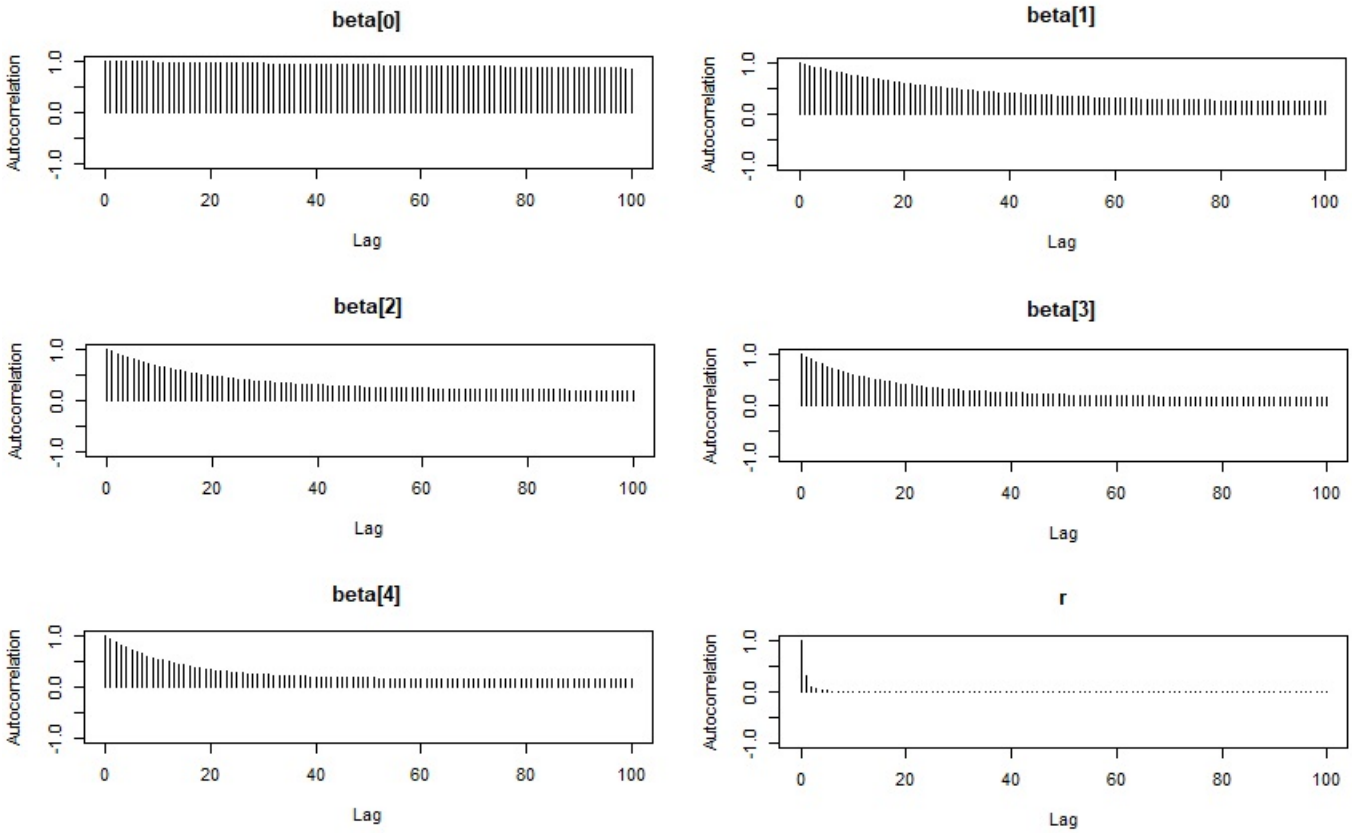


Table 5.13: Zero-Inflated Negative Binomial Heidelberg and Welch Diagnostics

	Stationarity and Half-width Test	Start Iteration	P-value	Mean	Half-width
Intercept	Passed	1	0.309	-1.862	2.85e-03
Bonus	Passed	120,001	0.327	-0.191	8.28e-05
Kilometres	Passed	120,001	0.753	0.114	1.77e-04
Make	Passed	120,001	0.571	-0.021	7.09e-05
Zone	Passed	120,001	0.256	-0.101	1.17e-04
r	Passed	1	0.996	11.201	2.40e-03

Table 5.14: Zero-Inflated Negative Binomial Posterior Results

	Mean	SD	Naïve SE	Time-Series SE
Intercept	-1.86180	0.044205	4.035e-05	1.455e-03
Bonus	-0.19105	0.005415	4.943e-06	4.027e-05
Kilometres	0.11435	0.009023	8.237e-06	8.639e-05
Make	-0.02099	0.005008	4.572e-06	3.472e-05
Zone	-0.10075	0.006830	6.235e-06	5.710e-05
r (shape)	11.20135	0.828431	7.563e-04	1.225e-03

Table 5.15: Zero-Inflated Negative Binomial Credible Intervals

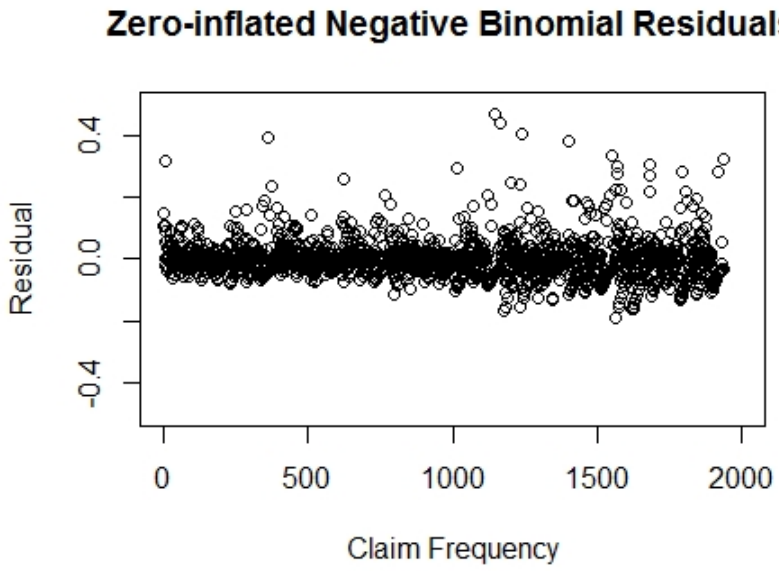
	2.5%	25%	50%	75%	97.5%
Intercept	-1.95083	-1.89085	-1.86174	-1.83236	-1.77568
Bonus	-0.20166	-0.19471	-0.19105	-0.18741	-0.18041
Kilometres	0.09684	0.10823	0.11434	0.12045	0.13212
Make	-0.03076	-0.02437	-0.02099	-0.01762	-0.01114
Zone	-0.11417	-0.10535	-0.10076	-0.09617	-0.08732
r (shape)	9.67489	10.62743	11.16675	11.73727	12.92690

supports the findings from the Zero-inflated Poisson: the data does not contain enough zero frequencies to warrant Zero-inflated models. The full results from the Zero-inflated Negative Binomial model are given by Table 7.8 in the Appendix. Table 5.15 in the Appendix also contains the credible intervals for the Zero-inflated Negative Binomial model. The randomness of the residuals about zero in Figure 5.12 are another indication of an appropriate model fit.

Table 5.16: Claim Frequency by Zone (Zero-Inflated Negative Binomial Results)

Code	Zone	Claim Frequency
1	Stockholm, Goteborg, Malmo with surroundings	0.09144302
2	Other large cities and surroundings	0.08267904
3	Small cities in northern Sweden	0.074755
4	Small cities in southern Sweden	0.06759041
5	Rural areas in northern Sweden	0.06108431
6	Rural areas in southern Sweden	0.0552554
7	Gotland	0.04807919

Figure 5.12: Residual Plot for Zero-Inflated Negative Binomial Model



5.1.5 Comparison

As it was shown in Section 5.1.1 and Section 5.1.3, both the Poisson and Negative Binomial models produced an appropriate fit for the claim frequency of the dataset. The DIC scores are therefore analysed in order to establish the best model. Since the DIC measures posterior predictive error, once the MCMC iterations converge, a fair comparison of different iteration lengths is possible. The DIC for the Negative Binomial model is 12% lower than the Poisson model. As a result of the over-dispersion, the Negative Binomial model for claim frequency is a better fit. The Zero-inflated models show no significant improvements. This is to be expected after the similar claim frequency results as shown in Section 5.1. These models are further compared in the expected compensation Table 5.23.

Table 5.17: DIC Criterion for Claim Frequency

Model	Mean Deviance	Penalty	Penalized Deviance
Poisson	10,237	5.221	10,242
Zero-inflated Poisson	10,240	10.34	10,250
Negative Binomial	9,023	6.3	9,029
Zero-inflated Negative Binomial	9,022	5.376	9,027

5.2 Claim Severity

5.2.1 Gamma Model

The Gamma model obtained after running 75,000 iterations is given by

$$E\left(\frac{\theta_i}{e_i}\right) = exp(8.261333 + 0.029702x_1 + 0.033309x_2 + 0.006689x_3 + 0.016532x_4). \quad (45)$$

This is the expected claim severity per claim made. It represents the posterior distributions for each variable in the Gamma MCMC sampler. For more information, see Table 5.19. The trace plots for the Gamma variables are all convergent. The MCMC iterations are further analysed to ensure convergence with the Heidelberg and Welch diagnostic. The lag plots are given in Figure 5.14.

The intercept β_0 has much higher autocorrelation than would be preferred. This implies that longer chains need to be run in order to insure convergence. Once again thinning is not used since we are more interested in the accuracy of the posterior estimates. The Heidelberg and Welch tests are both passed for all variables in the Gamma model as seen in Table 5.18. None of the iterations from the chain are discarded aside from the first 10% of the Make variable. The larger p-values for the intercept and zone indicate weak evidence against the null hypothesis. The null hypothesis is not rejected for all variables. This is a good indicator that the chains tend towards convergence. The credible intervals for the Gamma model are reasonable. They are discussed further in the analysis of claim severity differences. Credible intervals for claim severity are given in Table 5.20

Figure 5.13: Trace Plots and Posterior Density Plots for Gamma Iterations. 3 Chains With 75,000 Iterations Each

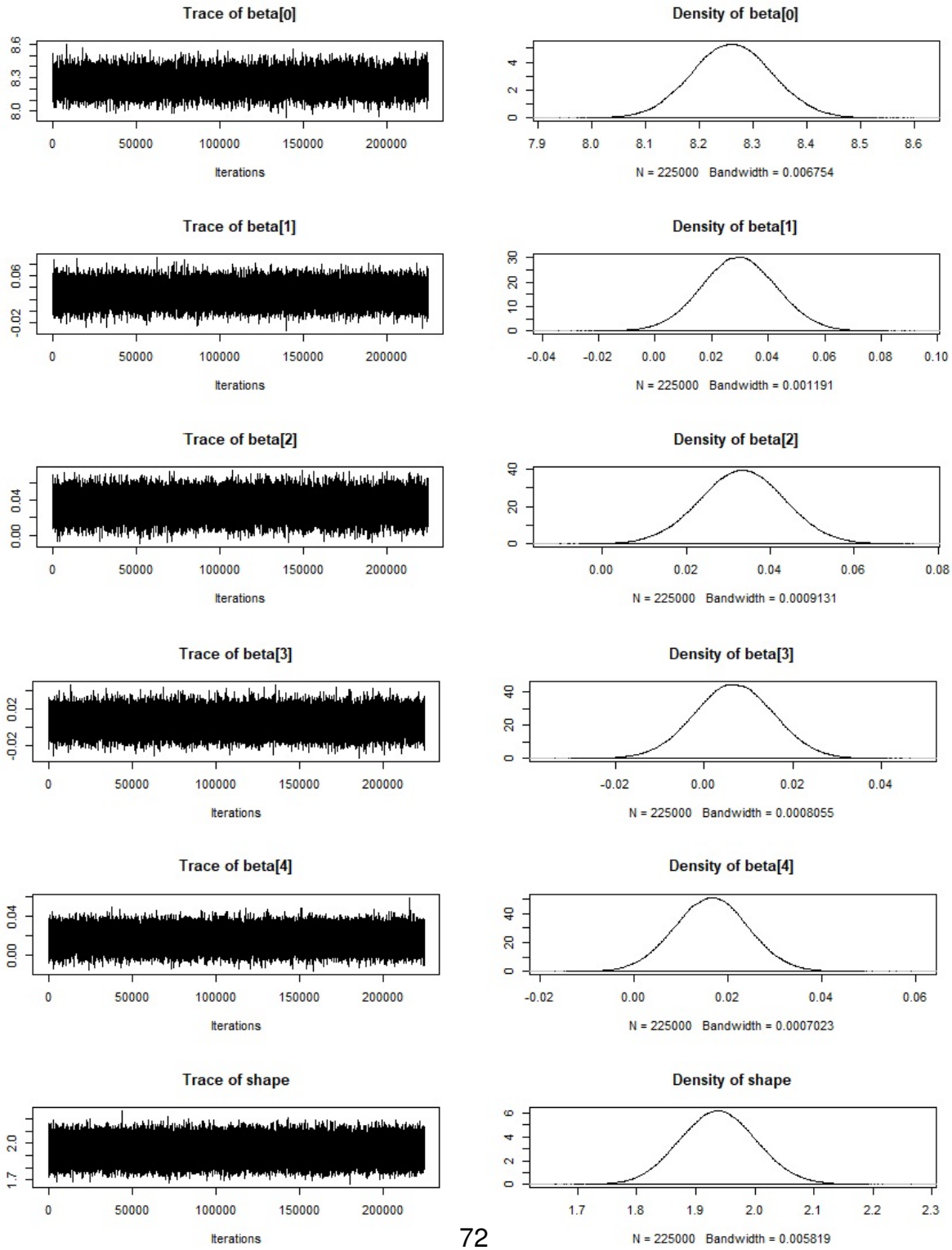


Figure 5.14: Autocorrelation Plot for Gamma Iterations

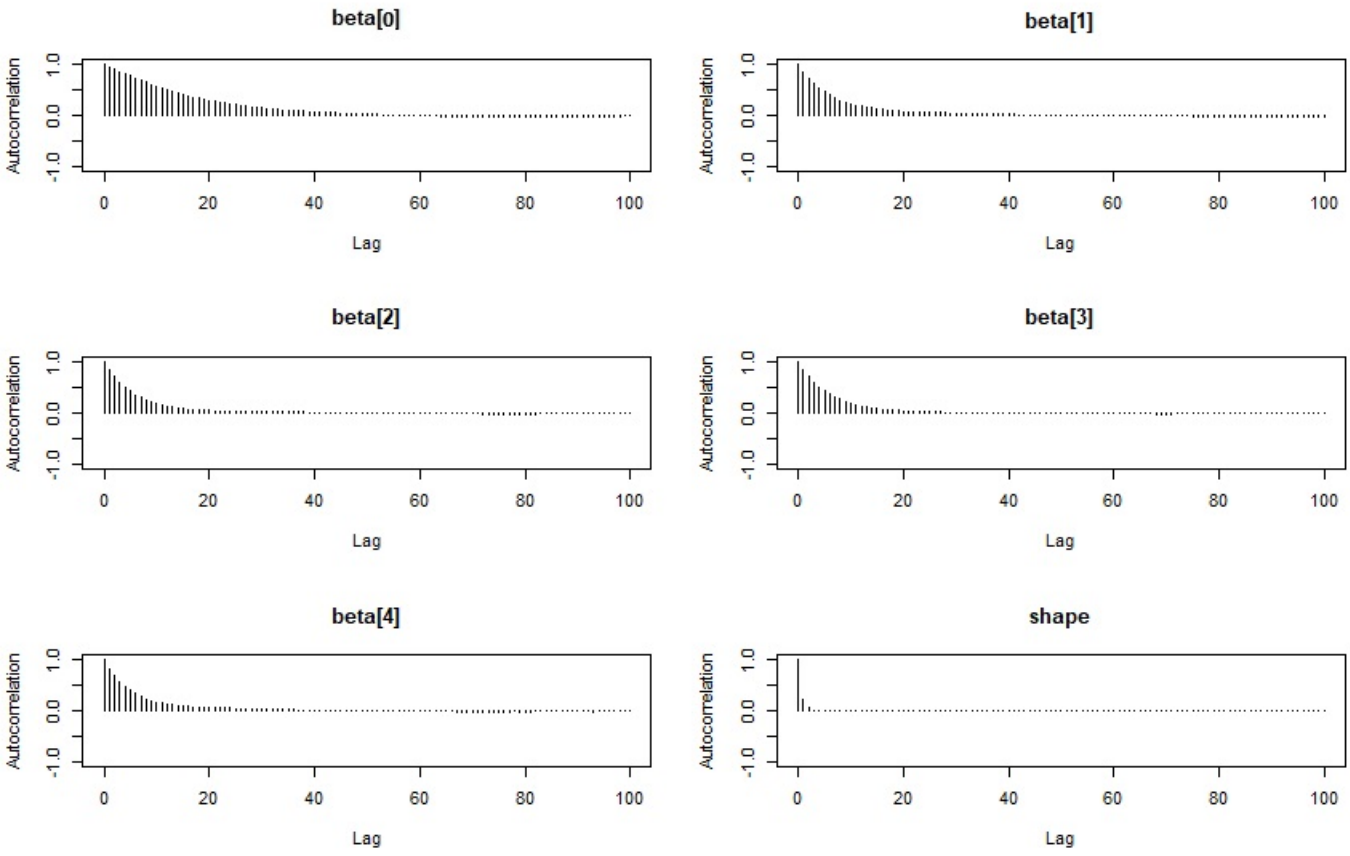


Table 5.18: Gamma Heidelberg and Welch Diagnostics

	Stationarity and Half-width Test	Start Iteration	P-value	Mean	Half-width
Intercept	Passed	1	0.636	8.26120	0.001841
Bonus	Passed	1	0.242	0.00668	0.000135
Kilometres	Passed	1	0.249	0.02967	0.000220
Make	Passed	22501	0.312	0.03328	0.000158
Zone	Passed	1	0.896	0.01656	0.000114
r	Passed	1	0.591	11.201	2.40e-03

in the Appendix. The DIC for the Gamma model is also given in Table 5.22 as well as the residuals in Figure 5.15. As seen in Table 5.21, the claim severity is decreasing when moving away from major cities. The claim severities follows the same trends as the claim frequencies, where the larger severity are in the more densely populated cities and smaller severities are in the rural parts of the country.

The Gamma model is run with different combinations of variables. The DIC scores for the different models was analysed. The model given above had the lowest DIC score. Therefore, this model is kept for calculating the expected compensation.

Table 5.19: Gamma Posterior Results

	Mean	SD	Naïve SE	Time-Series SE
Intercept	8.261333	0.074940	1.580e-04	9.177e-04
Bonus	0.006689	0.008937	1.884e-05	6.622e-05
Kilometres	0.029702	0.013218	2.787e-05	1.063e-04
Make	0.016532	0.007792	1.643e-05	5.922e-05
Zone	0.033309	0.010136	2.137e-05	7.382e-05
r (shape)	1.938478	0.064565	1.361e-04	1.733e-04

Table 5.20: Gamma Credible Intervals

	2.5%	25%	50%	75%	97.5%
Intercept	8.115721	8.2105063	8.260930	8.31152	8.40945
Bonus	-0.010914	0.0006828	0.006688	0.01273	0.02413
Kilometres	0.003820	0.0207980	0.029646	0.03859	0.05574
Make	0.001261	0.0112792	0.016543	0.02177	0.03185
Zone	0.013326	0.0265453	0.033327	0.04012	0.05317
r (shape)	1.814190	1.8944798	1.937981	1.98158	2.06713

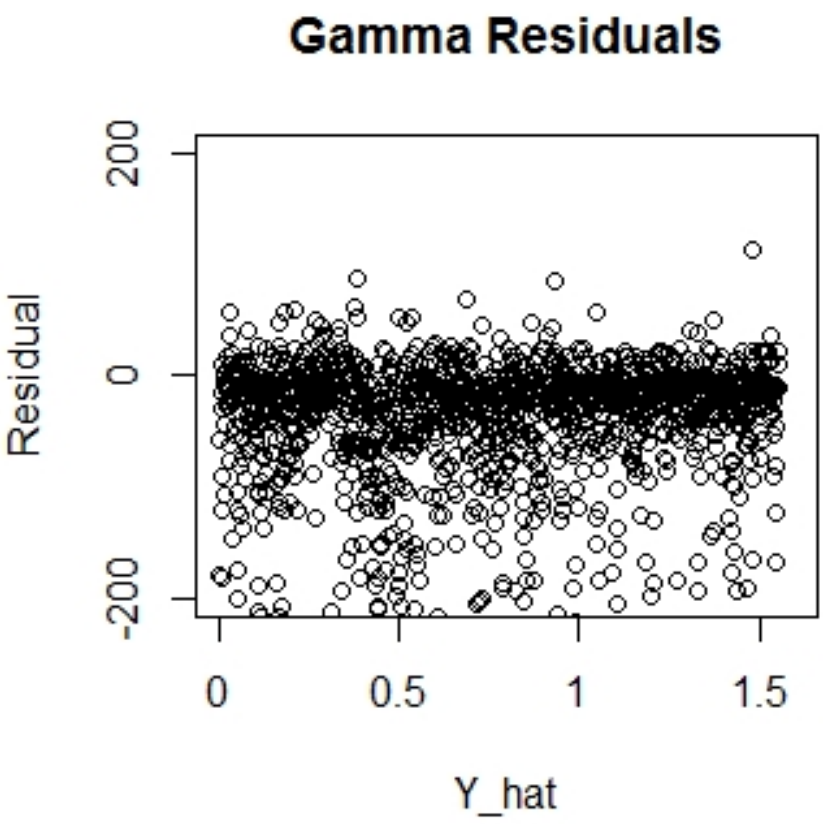
Table 5.21: Claim Severity by Zone (Gamma Results)

Code	Zone	Claim Severity
1	Stockholm, Goteborg, Malmo with surroundings	119,536.6
2	Other large cities and surroundings	147,291.5
3	Small cities in northern Sweden	109,075.4
4	Small cities in southern Sweden	102,699.2
5	Rural areas in northern Sweden	61,121.93
6	Rural areas in southern Sweden	47,180.94
7	Gotland	28,610.07

Table 5.22: DIC scores for the Gamma Model

Model	Mean Deviance	Penalty	Penalized Deviance
Gamma	35,550	3.934	35,554

Figure 5.15: Residual Plot for Gamma Model ($1e^3$)



5.3 Expected Compensation

Table 5.23 and Table 5.3 are comparable. The results for Table 5.23 support a higher expected compensation in metropolitan and larger cities, and smaller compensation in the rural country. The results from Table 5.3 support these findings. Table 5.3 also shows the positive correlation between kilometres driven and expected compensation. The results observed for metropolitan and large cities of the Negative Binomial results indicate a larger expected compensation compared to the Poisson. The smaller the expected compensation, the closer the results are for the models. This is consonant with the fact that over-dispersed data is not accounted for in a Poisson model. Therefore, the Negative Binomial model better accounts for the points with claims that are outside 2σ. This is reflected in both the claim frequency and the expected compensation results.

			Expected Compensation			
Code	Zone	KMs/Year	Poisson	ZIP	Negative Binomial	ZINB
1-1	Metropolitans	$<$1,000	7,703	7,708	8,483	8,488
1-2	Metropolitans	1,000-15,000	12,516	12,524	13,365	13,372
1-3	Metropolitans	15,000-20,000	9,357	9,362	9,688	9,691
1-4	Metropolitans	20,000-25,000	4,242	4,244	4,259	4,259
1-5	Metropolitans	$>$25,000	4,312	4,314	4,197	4,198

Continued on the next page

Code	Zone	KMs/Year	Poisson	ZIP	Negative Binomial	ZINB
2-1	Large Cities	<1,000	6,691	6,695	7,408	7,411
2-2	Large Cities	1,000-15,000	11,230	11,235	12,055	12,059
2-3	Large Cities	15,000-20,000	8,595	8,598	8,946	8,948
2-4	Large Cities	20,000-25,000	4,544	4,545	4,586	4,586
2-5	Large Cities	>25,000	4,374	4,375	4,280	4,280
3-1	Small cities Northern	<1,000	6,213	6,216	6,915	6,918
3-2	Small cities Northern	1,000-15,000	10,060	10,063	10,856	10,859
3-3	Small cities Northern	15,000-20,000	7,609	7,611	7,962	7,963
3-4	Small cities Northern	20,000-25,000	4,280	4,281	4,342	4,342
3-5	Small cities Northern	>25,000	3,985	3,985	3,920	3,920
4-1	Small cities Southern	<1,000	9,046	9,048	10,120	10,123
4-2	Small cities Southern	1,000-15,000	15,276	15,279	16,571	16,574
4-3	Small cities Southern	15,000-20,000	12,376	12,377	13,017	13,018
4-4	Small cities Southern	20,000-25,000	7,242	7,242	7,386	7,386
4-5	Small cities Southern	>25,000	6,069	6,069	6,002	6,001
5-1	Rural Northern	<1,000	1,930	1,930	2,170	2,171

Continued on the next page

Code	Zone	KMs/Year	Poisson	ZIP	Negative Binomial	ZINB
5-2	Rural Northern	1,000-15,000	3,154	3,154	3,439	3,440
5-3	Rural Northern	15,000-20,000	2,224	2,223	2,351	2,351
5-4	Rural Northern	20,000-25,000	1,267	1,266	1,299	1,298
5-5	Rural Northern	$>$25,000	1,503	1,502	1,494	1,493
6-1	Rural Southern	$<$1,000	3,157	3,156	3,569	3,569
6-2	Rural Southern	1,000-15,000	4,915	4,914	5,387	5,387
6-3	Rural Southern	15,000-20,000	4,372	4,372	4,647	4,647
6-4	Rural Southern	20,000-25,000	2,300	2,300	2,371	2,370
6-5	Rural Southern	$>$25,000	2,226	2,225	2,224	2,223
7-1	Gotland	$<$1,000	396	396	450	450
7-2	Gotland	1,000-15,000	477	477	526	526
7-3	Gotland	15,000-20,000	541	541	578	578
7-4	Gotland	20,000-25,000	390	390	402	402
7-5	Gotland	$>$25,000	462	462	463	463

Table 5.23: Expected Compensation by Zone

		Expected Compensation			
Code	Zone	Poisson	Zero-inflated Poisson	Negative Binomial	Zero-inflated Negative Binomial
1	Stockholm, Goteborg, Malmo with surroundings	10,638.0103	10,643.68231	10,927.31517	10,930.7877
2	Other large cities and surroundings	11,790.85838	11,795.37581	12,175.14632	12,177.91982
3	Small cities in northern Sweden	7,854.214143	7,856.045519	8,152.80805	8,153.931527
4	Small cities in southern Sweden	6,651.984314	6,652.536835	6,941.149316	6,941.481035
5	Rural areas in northern Sweden	3,558.032234	3,557.801193	3,733.739446	3,733.59092
6	Rural areas in southern Sweden	2,472.675689	2,472.139241	2,607.346605	2,607.001712
7	Gotland	1,294.539578	1,294.068943	1,375.8411	1,375.548991

6 Conclusion

The objective of this study was to compare and contrast the different claim frequency models in the industry and use the results, along with the results for the predicted claim severity, to produce expected compensation cost for third party insurance companies. The most common model over the past years has been the Poisson model primarily owing to its simplicity. It is used for a quick appraisal of a customer. In theory, the Negative Binomial is more accurate, but at the cost of a higher computational power. This makes a significant difference in the computational time when running the MCMC algorithm. Zero-inflated models are best utilized in the industry when zero valued observations are frequent. In this thesis, four kinds of models were used and compared to best fit the data.

The expected compensation cost was calculated by the frequency-severity method. That is, the claim frequency multiplied by the claim severity. Results from the claim frequency models were combined with the results from the claim severity. They are summarized below:

- **Claim Frequency**
  - Dependence between claim frequency and zone of policyholders
  - Comparison between Poisson and Negative Binomial and Zero-inflated models
  - The best results come from Negative Binomial models as per DIC scores
  - Claim frequency is the highest in the largest cities and lowest in rural places

- **Claim Severity**
  - Claim severity is given by zone and kilometres driven
  - Significant difference in claim severity from one zone to the next. With larger severity coming from large cities and less severe from rural places
- **Expected Compensation**
  - Highest cost in largest cities and lowest cost in rural countryside
  - Considered with the claim frequency and claim severity

According to these results, there is a significant difference in expected compensation by zone and kilometres driven. It is confirmed that pricing car insurance by zone and yearly kilometres driven is appropriate practice. The standard Poisson model is still accurate, however, depending on the data set and data size, the Negative Binomial or the Zero-inflated models might produce a better fit.

7 Future Work

It should be noted that independence between claim severity and inter-claim arrival times was assumed in this thesis. In continued studies, it would be worth while looking at a model with possible dependence between claim severity and inter-claim arrival times. While working and doing researching for this thesis, a few key points for further work were determined. These items include:

- Additional data for more accurate results. These could include variables such as:
  - Gender of Policyholder
  - Age of Policyholder
  - Driving History of Policyholder
  - Additional Years of Data
  - License of Policyholder
- Add and test for spatial factors
- Add and test for temporal factors
- Compare models on different data sets which have:
  - More Zero Claim Frequencies
  - Different levels of dispersion
  - Limited data size to compare MCMC conversion for the different models

References

[Brooks and Gelman, 1998] Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455.

[Cameron and Trivedi, 1986] Cameron, A. C. and Trivedi, P. K. (1986). Econometric models based on count data. comparisons and applications of some estimators and tests. *Journal of applied econometrics*, 1(1):29–53.

[Cameron and Trivedi, 2013] Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge university press.

[Casella and George, 1992] Casella, G. and George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174.

[Depaoli et al., 2016] Depaoli, S., Clifton, J. P., and Cobb, P. R. (2016). Just another gibbs sampler (jags) flexible software for mcmc implementation. *Journal of Educational and Behavioral Statistics*, 41(6):628–649.

[Embrechts et al., 2013] Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media.

[Gelfand et al., 2005] Gelfand, A. E., Banerjee, S., and Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics: The official journal of the International Environmetrics Society*, 16(5):465–479.

[Gelman et al., 2013] Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

[Geman and Geman, 1984] Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741.

[Gill, 2002] Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach*. Chapman and Hall/CRC.

[Gourieroux and Jasiak, 2001] Gourieroux, C. and Jasiak, J. (2001). Dynamic factor models. *Econometric Reviews*, 20(4):385–424.

[Greene, 1994] Greene, W. H. (1994). Accounting for excess zeros and sample selection in poisson and negative binomial regression models.

[Haberman and Renshaw, 1996] Haberman, S. and Renshaw, A. E. (1996). Generalized linear models and actuarial science. *The Statistician*, pages 407–436.

[Heidelberger and Welch, 1983] Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144.

[Kaas et al., 2008] Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). *Modern actuarial risk theory: using R*, volume 128. Springer Science & Business Media.

[Kruschke and Liddell, 2018] Kruschke, J. K. and Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic bulletin & review*, 25(1):155–177.

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

[Link and Eaton, 2012] Link, W. A. and Eaton, M. J. (2012). On thinning of chains in mcmc. *Methods in ecology and evolution*, 3(1):112–115.

[McCullagh and Nelder, 1989] McCullagh, P. and Nelder, J. A. (1989). Generalized linear models, vol. 37 of monographs on statistics and applied probability.

[Migon and Moura, 2005] Migon, H. S. and Moura, F. A. (2005). Hierarchical bayesian collective risk model: an application to health insurance. *Insurance: Mathematics and Economics*, 36(2):119–135.

[Nelder and Wedderburn, 1972] Nelder, J. A. and Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384.

[Panjer and Willmot, 1983] Panjer, H. H. and Willmot, G. E. (1983). Compound poisson models in actuarial risk theory. *Journal of Econometrics*, 23(1):63–76.

[Renshaw, 1994] Renshaw, A. E. (1994). Modelling the claims process in the presence of covariates. *ASTIN Bulletin: The Journal of the IAA*, 24(2):265–285.

[Shilane et al.,] Shilane, D., Evans, S. N., and Hubbard, A. E. Confidence intervals for negative binomial random variables of high dispersion. *The international journal of biostatistics*, 6(1).

[Silverman, 2018] Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

[Smith et al., 2007] Smith, B. J. et al. (2007). boa: an r package for mcmc output convergence assessment and posterior inference. *Journal of Statistical Software*, 21(11):1–37.

[Spiegelhalter et al., 1996] Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1996). Bugs 0.5: Bayesian inference using gibbs sampling manual (version ii). *MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK*, pages 1–59.

[Spiegelhalter et al., 2002] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

[Toft et al., 2007] Toft, N., Innocent, G. T., Gettinby, G., and Reid, S. W. (2007). Assessing the convergence of markov chain monte carlo methods: an example from evaluation of diagnostic tests in absence of a gold standard. *Preventive veterinary medicine*, 79(2-4):244–256.

Table 7.1: 95% Credible Interval for Claim Frequency (Poisson Model)

Code	Zone	Lower Bound	Upper Bound
1	Stockholm, Goteborg, Malmo with surroundings	0.07855981	0.1008392
2	Other large cities and surroundings	0.07011875	0.09140476
3	Small cities in northern Sweden	0.06258466	0.08285297
4	Small cities in southern Sweden	0.05586009	0.07510128
5	Rural areas in northern Sweden	0.0498232	0.06800349
6	Rural areas in southern Sweden	0.04450094	0.06170578
7	Gotland	0.03818582	0.05359728

Appendix

Table 7.2: Claim Frequency by Zone and Kilometres Driven per Year (Poisson Model)

Code	Zone	Kilometres/Year	Claim Frequency
1-1	Stockholm, Goteborg, Malmo with surroundings	$<$1,000	0.0651638
1-2	Stockholm, Goteborg, Malmo with surroundings	1,000-15,000	0.07535604
1-3	Stockholm, Goteborg, Malmo with surroundings	15,000-20,000	0.08714244

Continued on the next page

Code	Zone	Kilometres/Year	Claim Frequency
1-4	Stockholm, Goteborg, Malmo with surroundings	20,000-25,000	0.1007724
1-5	Stockholm, Goteborg, Malmo with surroundings	>25,000	0.1165341
2-1	Other large cities and surroundings	<1,000	0.05861579
2-2	Other large cities and surroundings	1,000-15,000	0.06778386
2-3	Other large cities and surroundings	15,000-20,000	0.0783859
2-4	Other large cities and surroundings	20,000-25,000	0.0906462
2-5	Other large cities and surroundings	>25,000	0.1048241
3-1	Small cities in northern Sweden	<1,000	0.05272576
3-2	Small cities in northern Sweden	1,000-15,000	0.06097257
3-3	Small cities in northern Sweden	15,000-20,000	0.07050926
3-4	Small cities in northern Sweden	20,000-25,000	0.08153758
3-5	Small cities in northern Sweden	>25,000	0.09429084
4-1	Small cities in southern Sweden	<1,000	0.04742759
4-2	Small cities in southern Sweden	1,000-15,000	0.05484571
4-3	Small cities in southern Sweden	15,000-20,000	0.06342411
4-4	Small cities in southern Sweden	20,000-25,000	0.07334425

Continued on the next page

Code	Zone	Kilometres/Year	Claim Frequency
4-5	Small cities in southern Sweden	>25,000	0.08481599
5-1	Rural areas in northern Sweden	<1,000	0.04266181
5-2	Rural areas in northern Sweden	1,000-15,000	0.04933452
5-3	Rural areas in northern Sweden	15,000-20,000	0.05705091
5-4	Rural areas in northern Sweden	20,000-25,000	0.06597422
5-5	Rural areas in northern Sweden	>25,000	0.07669898
6-1	Rural areas in southern Sweden	<1,000	0.03837492
6-2	Rural areas in southern Sweden	1,000-15,000	0.04437712
6-3	Rural areas in southern Sweden	15,000-20,000	0.05131813
6-4	Rural areas in southern Sweden	20,000-25,000	0.05934477
6-5	Rural areas in southern Sweden	>25,000	0.06862686
7-1	Gotland	<1,000	0.03415269
7-2	Gotland	1,000-15,000	0.03991787
7-3	Gotland	15,000-20,000	0.0461614
7-4	Gotland	20,000-25,000	0.05205858
7-5	Gotland	>25,000	0.05550082

Continued on the next page

Code	Zone	Kilometres/Year	Claim Frequency

Table 7.4: Claim Frequency by Zone and Kilometres Driven per Year (Zero-Inflated Poisson Model)

Code	Zone	Kilometres/Year	Claim Frequency
1-1	Stockholm, Goteborg, Malmo with surroundings	$<$1,000	0.06520899
1-2	Stockholm, Goteborg, Malmo with surroundings	1,000-15,000	0.07540302
1-3	Stockholm, Goteborg, Malmo with surroundings	15,000-20,000	0.08719067
1-4	Stockholm, Goteborg, Malmo with surroundings	20,000-25,000	0.1008211
1-5	Stockholm, Goteborg, Malmo with surroundings	$>$25,000	0.1165823
2-1	Other large cities and surroundings	$<$1,000	0.05864764
2-2	Other large cities and surroundings	1,000-15,000	0.06781594
2-3	Other large cities and surroundings	15,000-20,000	0.07841751
2-4	Other large cities and surroundings	20,000-25,000	0.09067641
2-5	Other large cities and surroundings	$>$25,000	0.1048517
3-1	Small cities in northern Sweden	$<$1,000	0.05274649
3-2	Small cities in northern Sweden	1,000-15,000	0.06099228

Continued on the next page

Code	Zone	Kilometres/Year	Claim Frequency
3-3	Small cities in northern Sweden	15,000-20,000	0.07052712
3-4	Small cities in northern Sweden	20,000-25,000	0.08155252
3-5	Small cities in northern Sweden	>25,000	0.09430152
4-1	Small cities in southern Sweden	<1,000	0.04743912
4-2	Small cities in southern Sweden	1,000-15,000	0.05485522
4-3	Small cities in southern Sweden	15,000-20,000	0.06343066
4-4	Small cities in southern Sweden	20,000-25,000	0.07334668
4-5	Small cities in southern Sweden	>25,000	0.08481287
5-1	Rural areas in northern Sweden	<1,000	0.04266578
5-2	Rural areas in northern Sweden	1,000-15,000	0.04933567
5-3	Rural areas in northern Sweden	15,000-20,000	0.05704824
5-4	Rural areas in northern Sweden	20,000-25,000	0.06596652
5-5	Rural areas in northern Sweden	>25,000	0.07668497
6-1	Rural areas in southern Sweden	<1,000	0.03837274
6-2	Rural areas in southern Sweden	1,000-15,000	0.0443715
6-3	Rural areas in southern Sweden	15,000-20,000	0.05130803

Continued on the next page

Code	Zone	Kilometres/Year	Claim Frequency
6-4	Rural areas in southern Sweden	20,000-25,000	0.05932894
6-5	Rural areas in southern Sweden	>25,000	0.06860376
7-1	Gotland	<1,000	0.0341454
7-2	Gotland	1,000-15,000	0.03990682
7-3	Gotland	15,000-20,000	0.0461454
7-4	Gotland	20,000-25,000	0.05203696
7-5	Gotland	>25,000	0.05547081

Table 7.6: Claim Frequency by Zone and Kilometres Driven per Year (Negative Binomial Model)

Code	Zone	Kilometres/Year	Claim Frequency
1-1	Stockholm, Goteborg, Malmo with surroundings	<1,000	0.07176465
1-2	Stockholm, Goteborg, Malmo with surroundings	1,000-15,000	0.08046819
1-3	Stockholm, Goteborg, Malmo with surroundings	15,000-20,000	0.09022729
1-4	Stockholm, Goteborg, Malmo with surroundings	20,000-25,000	0.10117
1-5	Stockholm, Goteborg, Malmo with surroundings	>25,000	0.1134398

Continued on the next page

Code	Zone	Kilometres/Year	Claim Frequency
2-1	Other large cities and surroundings	<1,000	0.06489249
2-2	Other large cities and surroundings	1,000-15,000	0.07276259
2-3	Other large cities and surroundings	15,000-20,000	0.08158717
2-4	Other large cities and surroundings	20,000-25,000	0.09148198
2-5	Other large cities and surroundings	>25,000	0.1025768
3-1	Small cities in northern Sweden	<1,000	0.05867842
3-2	Small cities in northern Sweden	1,000-15,000	0.06579488
3-3	Small cities in northern Sweden	15,000-20,000	0.07377441
3-4	Small cities in northern Sweden	20,000-25,000	0.0827217
3-5	Small cities in northern Sweden	>25,000	0.09275411
4-1	Small cities in southern Sweden	<1,000	0.0530594
4-2	Small cities in southern Sweden	1,000-15,000	0.05949439
4-3	Small cities in southern Sweden	15,000-20,000	0.06670981
4-4	Small cities in southern Sweden	20,000-25,000	0.07480031
4-5	Small cities in southern Sweden	>25,000	0.08387201
5-1	Rural areas in northern Sweden	<1,000	0.04797845

Continued on the next page

Code	Zone	Kilometres/Year	Claim Frequency
5-2	Rural areas in northern Sweden	1,000-15,000	0.05379723
5-3	Rural areas in northern Sweden	15,000-20,000	0.0603217
5-4	Rural areas in northern Sweden	20,000-25,000	0.06763746
5-5	Rural areas in northern Sweden	$>$25,000	0.07624006
6-1	Rural areas in southern Sweden	$<$1,000	0.04338405
6-2	Rural areas in southern Sweden	1,000-15,000	0.04864563
6-3	Rural areas in southern Sweden	15,000-20,000	0.05454532
6-4	Rural areas in southern Sweden	20,000-25,000	0.06116053
6-5	Rural areas in southern Sweden	$>$25,000	0.06857801
7-1	Gotland	$<$1,000	0.03881915
7-2	Gotland	1,000-15,000	0.04398734
7-3	Gotland	15,000-20,000	0.04932209
7-4	Gotland	20,000-25,000	0.05367565
7-5	Gotland	$>$25,000	0.05566228

Table 7.3: Credible intervals for claim frequency by zone (Zero-Inflated Poisson)

Code	Zone	Lower Bound	Upper Bound
1	Stockholm, Goteborg, Malmo with surroundings	0.07890548	0.1006524
2	Other large cities and surroundings	0.0704139	0.09122905
3	Small cities in northern Sweden	0.06283616	0.08268791
4	Small cities in southern Sweden	0.05607391	0.07494641
5	Rural areas in northern Sweden	0.05000433	0.06785885
6	Rural areas in southern Sweden	0.0446543	0.06156991
7	Gotland	0.03830817	0.05347759

Table 7.5: 95% Credible Interval for Negative Binomial Posterior Claim Frequency

Code	Zone	Lower Bound	Upper Bound
1	Stockholm, Goteborg, Malmo with surroundings	0.07175434	0.1165637
2	Other large cities and surroundings	0.06400994	0.1068267
3	Small cities in northern Sweden	0.05710138	0.09790305
4	Small cities in southern Sweden	0.05093847	0.08972485
5	Rural areas in northern Sweden	0.04543404	0.08216586
6	Rural areas in southern Sweden	0.04053632	0.07536084
7	Gotland	0.03490089	0.06627785

Table 7.7: 95% Credible Interval for Zero-Inflated Negative Binomial Claim Frequency

Code	Zone	Lower Bound	Upper Bound
1	Stockholm, Goteborg, Malmo with surroundings	0.07224435	0.1157341
2	Other large cities and surroundings	0.06444964	0.1060569
3	Small cities in northern Sweden	0.05749593	0.09718881
4	Small cities in southern Sweden	0.05129248	0.08906225
5	Rural areas in northern Sweden	0.0457512	0.08155234
6	Rural areas in southern Sweden	0.0408213	0.07479085
7	Gotland	0.03514252	0.06577356

Table 7.8: Claim Frequency by Zone and Kilometres Driven per Year (Zero-Inflated Negative Binomial Model)

Code	Zone	Kilometres/Year	Claim Frequency
1-1	Stockholm, Goteborg, Malmo with surroundings	<1,000	0.07180665
1-2	Stockholm, Goteborg, Malmo with surroundings	1,000-15,000	0.08050562
1-3	Stockholm, Goteborg, Malmo with surroundings	15,000-20,000	0.09025843
1-4	Stockholm, Goteborg, Malmo with surroundings	20,000-25,000	0.1011927
1-5	Stockholm, Goteborg, Malmo with surroundings	>25,000	0.1134517
2-1	Other large cities and surroundings	<1,000	0.06492463
2-2	Other large cities and surroundings	1,000-15,000	0.07278989
2-3	Other large cities and surroundings	15,000-20,000	0.08160798
2-4	Other large cities and surroundings	20,000-25,000	0.09149433
2-5	Other large cities and surroundings	>25,000	0.1025784
3-1	Small cities in northern Sweden	<1,000	0.05870219
3-2	Small cities in northern Sweden	1,000-15,000	0.06581363
3-3	Small cities in northern Sweden	15,000-20,000	0.07378659
3-4	Small cities in northern Sweden	20,000-25,000	0.08272543

Continued on the next page

Code	Zone	Kilometres/Year	Claim Frequency
3-5	Small cities in northern Sweden	>25,000	0.09274715
4-1	Small cities in southern Sweden	<1,000	0.05307612
4-2	Small cities in southern Sweden	1,000-15,000	0.05950599
4-3	Small cities in southern Sweden	15,000-20,000	0.06671481
4-4	Small cities in southern Sweden	20,000-25,000	0.07479694
4-5	Small cities in southern Sweden	>25,000	0.08385818
5-1	Rural areas in northern Sweden	<1,000	0.04798925
5-2	Rural areas in northern Sweden	1000-15,000	0.05380288
5-3	Rural areas in northern Sweden	15,000-20,000	0.0603208
5-4	Rural areas in northern Sweden	20,000-25,000	0.06762833
5-5	Rural areas in northern Sweden	>25,000	0.07622086
6-1	Rural areas in southern Sweden	<1,000	0.04338991
6-2	Rural areas in southern Sweden	1,000-15,000	0.04864636
6-3	Rural areas in southern Sweden	15,000-20,000	0.0545396
6-4	Rural areas in southern Sweden	20,000-25,000	0.06114677
6-5	Rural areas in southern Sweden	>25,000	0.06855436

Continued on the next page

Table 7.9: 95% Credible Interval for Claim Severity (Gamma Model)

Code	Zone	Lower Bound	Upper Bound
1	Stockholm, Goteborg, Malmo with surroundings	81,885.76	175,745.3
2	Other large cities and surroundings	98,292.15	222,025
3	Small cities in northern Sweden	72,001.4	166,305.4
4	Small cities in southern Sweden	66,825.96	158,797.2
5	Rural areas in northern Sweden	39,270.42	95,624.62
6	Rural areas in southern Sweden	28,292.5	79,124.09
7	Gotland	18,623.28	44,172.04

Code	Zone	Kilometres/Year	Claim Frequency
7-1	Gotland	$<$1,000	0.03882071
7-2	Gotland	1,000-15,000	0.04398405
7-3	Gotland	15,000-20,000	0.04931247
7-4	Gotland	20,000-25,000	0.05365848
7-5	Gotland	$>$25,000	0.05563551

The code for this paper was ran using R. The packaged used to run the MCMC iterations was Jags. It was introduced by Martyn Plummer as an open-source program for analysis and statistical inference of Bayesian hierarchical models [Depaoli et al., 2016]. **R Code for Poisson Program**

```
library(rjags)
df <- read.csv(file="c:/Users/Owner/Documents/insurance.csv",
   ↪ header=TRUE)
attach(df)
data = as.list(df)
Y <- cbind(df$Claims)
length(df$Kilometres)
X <- cbind(df$Kilometres, df$Zone, df$Bonus, df$Make)
X1 <- cbind(df$Kilometres, df$Zone, df$Bonus, df$Make)
par(mfrow=c(2,2))
df.2 = cbind(df$Kilometres, df$Zone, df$Bonus, df$Make, df$
   ↪ Insured, df$Claims)
pairs(df.2)
library(rjags)


model.sim = "␣model␣{
logInsured␣=␣log(Insured)
```

```
for␣(i␣in␣1:length(Claims)){
Claims[i]␣~␣dpois(lam[i])
log(lam[i])␣=␣intercept␣+␣b_Kilometres*Kilometres[i]␣+␣b_Zone*
   ↪ Zone[i]␣+␣b_Bonus*Bonus[i]␣+␣b_Make*Make[i]␣+␣1*logInsured[
   ↪ i]
res[i]␣=␣Claims[i]
}
intercept␣~␣dnorm(-1.89554,1/1e6)
b_Kilometres␣~␣dnorm(0.14531,1/1e4)
b_Zone␣~␣dnorm(-0.10604,1/1e4)
b_Bonus␣~␣dnorm(-0.19694,1/1e4)
b_Make␣~␣dnorm(-0.03695,1/1e4)
}"
set.seed(102)
data = as.list(df)
param = c("intercept","b_Kilometres","b_Zone","b_Bonus","b_Make")
model = jags.model(textConnection(model.sim),
data=data,
n.chains=3,
n.adapt = 1000)
update(model,1000)
```

```
inits = function(){list(intercept=rlnorm(1),b_Kilometres=rlnorm
   ↪ (1),b_Zone=rlnorm(1),b_Bonus=rlnorm(1),b_Make=rlnorm(1))}
traceplot(model.sim)
print(model.sim,dig =3)
hist(exp(model.sim$sims.list$intercept), xlab="Expected␣count",
   ↪ xlim = c(0,25), breaks=20)
model.sim = coda.samples(model=model,
variable.names = param,
n.iter = 30000,
thin = 2)


model.csim = as.mcmc(do.call(rbind,model.sim))
plot(model.csim)
summary(model.sim)
effectiveSize(model.sim)
raftery.diag(model.sim)
gelman.diag(model.sim, confidence = 0.95, transform = FALSE)
print(model.sim$mean, dig = 3)
autocorr.diag(model.sim)
autocorr.plot(model.sim, lag.max = 50)
effectiveSize(model.sim)
```

```
dic.poisson = dic.samples(model,n.iter = 1e3)

dic.poisson
```

R Code for Negative Binomial Program

```
mod.string2 = "␣model␣{

for␣(i␣in␣1:n){

mu[i]␣=␣beta[1]␣+␣beta[2]*Kil[i]␣+␣beta[3]*Zone[i]␣+␣beta[4]*

   ↪ Bonus[i]␣+␣beta[5]*Make[i]␣+␣1*Insu[i]

y[i]␣~␣dnegbin(p[i],alpha)

lambda[i]␣=␣exp(mu[i])

p[i]␣=␣alpha/(alpha+lambda[i])

}



beta[1:5]␣~␣dmnorm(b0[1:5],B0[,])

alpha␣~␣dunif(0,50)

}"

datanb = list(n = dim(df)[1],

Kil = df$Kilometres,

Zone = df$Zone,

Bonus = df$Bonus,

Make = df$Make,

y = df$Claims,
```

```
Insu = log(df$Insured),
b0=rep(0,5),
B0=diag(.0001,5))
writeLines(mod.string2, con="negbin1.bug")
set.seed(87)
require(rjags)
foo <- jags.model(file = "negbin1.bug",
data=datanb,
n.chains = 3,
n.adapt = 3000)

update(foo, 5000)
negbin <- coda.samples(foo,
variable.names=c("beta","alpha"),
n.iter=75000)

xyplot(negbin)
summary(negbin)
plot(negbin)
autocorr.diag(negbin)
autocorr.plot(negbin, lag.max = 100)
```

```
effectiveSize(negbin)

dic.negbin = dic.samples(foo,n.iter = 1e3)

dic.negbin
```

R Code for Zero-Inflated Poisson Program

```
zip = "model␣␣{

for(i␣in␣1:n){

y[i]␣~␣dpois(lambda.hacked[i])

lambda.hacked[i]␣=␣lambda[i]*(1-zero[i])␣+␣1e010*zero[i]

lambda[i]␣=␣exp(mu.count[i])

mu.count[i]␣=␣inprod(beta[],X[i,])␣+␣Insu[i]



zero[i]␣~␣dbern(pi[i])

pi[i]␣=␣ilogit(mu.binary[i])

mu.binary[i]␣=␣inprod(alpha[],X[i,])

}

beta␣~␣dmnorm(b0,B0)

alpha␣~␣dmnorm(a0,A0)

}"

set.seed(93)

dataZIP = list(n = dim(df)[1],

X = cbind(1,df$Kilometres,df$Zone,df$Bonus,df$Make),
```

```
y = df$Claims,
Insu = log(df$Insured),
b0=rep(0,5),
B0=diag(.0001,5),
a0=rep(0,5),
A0=diag(.0001,5))
writeLines(zip, con="zipoisson.bug")
zipi <- jags.model(file="zipoisson.bug",
data=dataZIP,
n.chains = 3,
n.adapt = 3000)

update(zipi, 5000)

ZIP<- coda.samples(zipi,
variable.names=c("beta", "alpha"),
n.iter=300000)
options(scipen = 999)
ZIpoisson = as.mcmc(do.call(rbind,ZIP))
plot(ZIpoisson)
summary(ZIP)
```

```
effectiveSize(ZIP)

raftery.diag(ZIP)

gelman.diag(ZIP, confidence = 0.95, transform = FALSE)



dic.zip = dic.samples(zipi,n.iter = 1e3)

dic.zip
```

R Code for Zero-Inflated Negative Binomial Program

```
ZINB = "model{

for (i in 1:n){

mu[i] = inprod(beta[],X[i,]) + Insu[i]

y[i] ~ dnegbin(p[i],r)

lambda[i] = exp(mu[i])

p[i] = (r/(r+lambda[i]))*(1-zero[i]) - 1e-10*zero[i]



zero[i] ~ dbern(pi[i])

pi[i] = ilogit(mu.binary[i])

mu.binary[i] = inprod(alpha[],X[i,])

}

beta[1:5] ~ dmnorm(b0[1:5],B0[,])

alpha[1:5] ~ dmnorm(a0[1:5],A0[,])

r ~ dunif(0,50)
```

```
}"
set.seed(93)
dataZINB = list(n = dim(df)[1],
X = cbind(1,df$Kilometres,df$Zone,df$Bonus,df$Make),
y = df$Claims,
Insu = log(df$Insured),
b0=rep(0,5),
B0=diag(.0001,5),
a0=rep(0,5),
A0=diag(.0001,5))
writeLines(ZINB, con="zinegbin.bug")
zinegbin <- jags.model(file="zinegbin.bug",
data=dataZINB,
n.chains = 3,
n.adapt = 3000)
update(zinegbin, 5000)
ZINEGBIN<- coda.samples(zinegbin,
variable.names=c("beta", "alpha", "r"),
n.iter=400000)
ZINB = as.mcmc(do.call(rbind,ZINEGBIN))
summary(ZINEGBIN)
```

```
plot(ZINEGBIN)

autocorr.diag(ZINEGBIN)

autocorr.plot(ZINEGBIN, lag.max = 100)

effectiveSize(ZINEGBIN)

dicZINB = dic.samples(zinegbin,n.iter = 1e3)

dicZINB
```

R Code for Gamma Program

```
Gamma.data2 = list(n = dim(dfpay2)[1],

Cl = dfpay2$Claims,

Y = dfpay2$Payment,

Kil = dfpay2$Kilometres,

Make = dfpay2$Make,

Bonus = dfpay2$Bonus,

Zone = dfpay2$Zone)

Gamma2 = "model{

for(j␣in␣1:5){beta[j]␣~␣dnorm(0.0,␣0.001)}

shape␣~␣dgamma(1,1)



for(i␣in␣1:n)␣{

mu[i]␣=␣log(Cl[i])*1␣+␣␣beta[1]␣+␣beta[2]*Kil[i]␣+␣beta[3]*Zone[i
   ↪ ]␣+␣beta[4]*Bonus[i]␣+␣beta[5]*Make[i]
```

```
Y[i] ~ dgamma(shape,shape/exp(mu[i]))
}
}"
Gamma.inits2 = function(){list("beta"=rep(0.001,5))}
Gamma.params2 = c("shape", paste("beta[",i=1:5,"]",sep=""))
Gamma2 = jags.model(textConnection(Gamma2),
data = Gamma.data2,
inits=Gamma.inits2,
n.chains = 3,
n.adapt = 5000)
update(Gamma2, 5000)
Gamma.model2 = coda.samples(model = Gamma2, variable.names =
   ↪ Gamma.params2, n.iter = 75000)
Gamma.model.csim2 = as.mcmc(do.call(rbind,Gamma.model2))
summary(Gamma.model2)
plot(Gamma.model.csim2)
autocorr.diag(Gamma2)
autocorr.plot(Gamma2, lag.max = 100)
effectiveSize(Gamma2)
dic.Gamma2 = dic.samples(foo,n.iter = 1e3)
dic.Gamma2
```