Analysis of claim frequency and claim size using the Bayesian approach

Erna Sigurgeirsdóttir

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science
University of Iceland
2010

ANALYSIS OF CLAIM FREQUENCY AND CLAIM SIZE USING THE BAYESIAN APPROACH

Erna Sigurgeirsdóttir

60 ECTS thesis submitted in partial fulfillment of a
Magister Scientiarum degree in Industrial Engineering

Advisor
Birgir Hrafnkelsson
Ólafur Pétur Pálsson

Faculty Representative
Guðmundur Freyr Úlfarsson

Faculty of Industrial Engineering, Mechanical Engineering and
Computer Science
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, May 2010

Analysis of claim frequency and claim size using the Bayesian approach
Analysis of insurance data using Bayesian approach
60 ECTS thesis submitted in partial fulfillment of a M.Sc. degree in Industrial Engineering

Faculty of Industrial Engineering, Mechanical Engineering and
Computer Science
School of Engineering and Natural Sciences
University of Iceland
Hjarðarhagi 2-6
107, Reykjavík, Reykjavik
Iceland

Telephone: 525 4000

Bibliographic information:
Erna Sigurgeirsdóttir, 2010, Analysis of claim frequency and claim size using the Bayesian approach, M.Sc. thesis, Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland.

Printing: Háskólaprent, Fálkagata 2, 107 Reykjavík
Reykjavik, Iceland, May 2010

Abstract

The aim of this thesis is to find the expected total compensation cost the insurance company (VÍS) needs to pay due to car accidents. Information are based on postal codes and the compensation cost is examined for each postal code in the country using a Poisson model and a gamma model, respectively. One of the main goals was to evaluate whether the residence of policyholders influence the result, by using spatial factor, one for each postal code. Another goal was to see if the results are improved by having temporal factor included in the model. Also, to improve the spatial factor, postal codes were categorized into postal category matrix where each category describes the type of region.

To examine the expected total compensation cost it is necessary to find the expected frequency of claims per policy year and the expected total claim size per policy year. These factors are examined per each postal code in the country. The expected compensation cost is calculated from the combination of those two models.

Data for mandatory liability insurances for vehicles are divided into three main categories, property loss, bodily injury and drivers accident insurance. In this study only results for property loss are given for one risk category over 16 year period.

The main result is that the claim frequency per policy year depends on the residence of policyholders. The number of claims are higher in the capital area than in the countryside. The spatial factor is very effective but the temporal factor influences much less. Nonetheless, it was decided to include the temporal factor in the model since it did not make the model less qualified. Also, there is a possibility of years with higher number of claims which would support the inclusion of the temporal factor.

Results for the total claim size show that the expected claim sizes are independent of the the residence of policyholders. The total claim size is similar no matter which part of the country is examined. Originally the model included both temporal factor and spatial factor but the spatial factor was taken out of the model since it did not influence the results. Nonetheless there are always some spatial effects from the postal category matrix. The temporal factor on the other hand has a substantial influence on the total claim size even though the data has been brought to present worth.

The expected total compensation cost the insurance company needs to pay is in context to the earlier mentioned results. The cost is highest in Reykjavík and the neighborhood of Reykjavík, gets lower when the countryside is examined but raises a bit at and around large urban regions. According to these results, it is justified that pricing of insurance would depend on the residence of policyholders.

Ágrip

Markmið þessa verkefnis er að finna út væntanlegan heildarkostnað sem tryggingafélagið (VÍS) þarf að greiða í bætur vegna bílslysa. Upplýsingar eru byggðar á póstnúmerum og er kostnaðurinn skoðaður fyrir hvert póstnúmer landsins. Hugmyndin er að meta hvort lögheimili vátryggingartaka hafi áhrif á niðurstöðuna með því að bæta landfræðilegum þætti í líkanið. Einnig var skoðað hvort bæta megi niðurstöður með því að hafa tímaþátt í líkaninu. Póstnúmerin voru flokkuð í stærri póstnúmeraflokka til að auka gæði landfræðilega þáttarins.

Til að skoða væntanlegan kostnað þarf fyrst að finna væntanlegan fjölda tjóna á hvert skírteinisár og væntanlega heildarstærð tjóna á hvert skírteinisár. Þessir tveir þættir eru skoðaðir fyrir hvert póstnúmer með því að nota Poisson líkan fyrir væntanlegan fjölda og gamma líkan fyrir heildarstærð. Væntanlegur kostnaður er reiknaður út frá sameiningu á þessum tveim líkönum.

Gögnum fyrir lögboðnar ábyrgðartryggingar vegna ökutækja er skipt í þrennt, munatjón, líkamstjón og slysatrygging ökumanns og eiganda. Aðeins eru gefnar niðurstöður fyrir munatjón í þessari ritgerð og fyrir einn áhættuflokk en gögnin ná yfir 16 ára tímabil.

Helstu niðurstöður eru að fjöldi tjóna á hvert skírteinisár er háður lögheimili vátryggingartaka. Fjöldi tjóna er meiri á höfuðborgarsvæðinu en á landsbyggðinni. Landfræðilegi þátturinn er áhrifamikill en tímaþátturinn hefur mun minni áhrif. Engu að síður var ákveðið að hafa tímaþáttinn í líkaninu þar sem það skerti ekki gæði líkansins. Einnig er möguleiki á tjónaþyngri árum sem styður það að nota tímaþáttinn.

Niðurstöður fyrir heildarstærð tjóna eru að lögheimili vátryggingartaka hefur lítil áhrif á heildar stærð tjóna. Póstnúmeraþátturinn hefur lítil áhrif, en engu að síður hafa póstnúmeraflokkarnir áhrif á niðurstöðuna. Heildarstærð tjóna er svipuð sama á hvaða landshluta er horft. Í upphafi innihélt líkanið bæði tímaþátt og landfræðilegan þátt en landfræðilegi þátturinn var tekinn út úr líkaninu þar sem hann hafði mjög lítil áhrif. Tímaþátturinn hefur aftur á móti mikil áhrif þrátt fyrir að gögnin hafi verið núvirt.

Væntanlegur heildarkostnaður sem tryggingafélagið þarf að greiða í bætur er í samræmi við niðurstöður sem búið er að greina frá. Kostnaðurinn er mestur í Reykjavík og nágrenni, lækkar eftir því sem kemur út á landsbyggðina en hækkar þó aðeins í og í kringum stærri þéttbýli. Kosnaðurinn er minnstur í dreifbýlum og minni þéttbýlum. Samkvæmt þessum niðurstöðum er réttlætanlegt að tryggingar séu verðlagðar eftir lögheimili vátryggingartaka.

Preface

This M.Sc. project was carried out at the Faculty of Industrial Engineering, Mechanical Engineering and Computer Science at the University of Iceland.

Special thanks goes to my supervisor, Associate Research Professor Birgir Hrafnkelsson at the Science Institute of University of Iceland, for all his time, guidance, encouragement and support throughout the study. I would also like to thank my co-advisor Professor Ólafur Pétur Pálsson at the Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, for his help during the writing of this thesis.

It́m very grateful for the co-operation with Friðrik Bragason and Ragnar Ragnarsson at VÍS and also I want to thank VÍS for their financial support during the study.

Furthermore, I would like to thank my friend Oddgeir Guðmundsson for technical support. Finally, I want to thank may family, friends and my partner, Stefán Elfar, for their moral support and encouragement during the study and also my loving son, Arnar Geir for his inspiration.

Reykjavík, May 2010.

Erna Sigurgeirsdóttir

Contents

List of Figures

List of Tables

Nomenclature

α Parameter in the gamma model

α_0 Parameter for gamma prior distribution

$\alpha_{\phi 1}$ and $\alpha_{\phi 2}$ Parameters in the prior distribution for ϕ_1 and ϕ_2 in the Poisson model

$\alpha_{\zeta 1}$ and $\alpha_{\zeta 2}$ Parameters in the prior distribution for ζ_1 and ζ_2 in the gamma model

β Coefficient for postal categories in the Poisson model

β_0 Parameter for gamma prior distribution

$\beta_{\phi 1}$ and $\beta_{\phi 2}$ Parameters in the prior distribution for ϕ_1 and ϕ_1 in the Poisson model

$\beta_{\zeta 1}$ and $\beta_{\zeta 2}$ Parameters in the prior distribution for ζ_1 and ζ_2 in the gamma model

η Coefficient for postal categories in the gamma model

κ_1^2 and κ_2^2 Hyperparameters in the gamma model

μ Mean

ν_r Parameter relative to σ_r^2 in the gamma model

$\nu_{\kappa 1}$ and $\nu_{\kappa 2}$ Parameters in the prior distribution for κ_1^2 and κ_2^2 in the gamma model

$\nu_{\tau 1}$ and $\nu_{\tau 2}$ Parameters in the prior distribution for τ_1^2 and τ_2^2 in the Poisson model

ν_ε Parameter in the prior distribution for σ_ε^2 in the Poisson model

ϕ_1 and ϕ_2 Hyperparameters in the Poisson model

$\sigma^2_{\varepsilon,it}$ Variance constant in the Poisson model

σ^2_{it} Variance in the gamma model

σ^2_r Hyperparameter in the gamma model

τ^2_1 and τ^2_2 Hyperparameters in the Poisson model

θ_{it} Average number of claims in region i and at year t

ε_{it} Error term in the Poisson model

ζ_1 and ζ_2 Hyperparameters in the gamma model

$a_{1,t}$ Temporal factor in the Poisson model

$a_{2,i}$ Regional factor in the Poisson model

C Neighborhood matrix based on the neighborhood matrix H

$d_{1,t}$ Temporal factor in the gamma model

$d_{2,i}$ Regional factor in the gamma model

e_{it} Number of policy-years in region i and at year t

H Neighborhood matrix

M Diagonal matrix based to the neighborhood matrix H

N_{it} Number of claim in region i and at year t

S^2_r Parameter relative to σ^2_r in the gamma model

$S^2_{\kappa 1}$ and $S^2_{\kappa 2}$ Parameters in the prior distribution for κ^2_2 in the gamma model

$S_{\tau 1}^2$ and $S_{\tau 2}^2$ Parameters in the prior distribution for τ_1^2 and τ_2^2 in the Poisson model

S_ε^2 Parameter in the prior distribution for σ_ε^2 in the Poisson model

S_{it} Total size of claim in region i and at year t

W Unknown cost for individual claim

x_i^T Covariate matrix for postal categories

1. Introduction

The data set in this study comes from the Icelandic insurance company VÍS or Vátryggingafélag Íslands hf.
Large part of cost for insurance companies are compensations because of traffic accidents. Mandatory liability insurances for vehicles are divided into three main categories,

i) Property loss

ii) Bodily injury

iii) Drivers accident insurance

The risk premium charged to policyholders is determined on the basis of the claims that the company expects during the period of policy. Therefore it is important for the insurance company to have an estimate of the expected compensation for all categories. To obtain this estimate is it is necessary to examine claim frequency and claim size which are the quantities behind expected compensation cost. Interest existed at the insurance company to examine this cost down to each postal code in the country and see if there is a spatial trend in the claims. Today, the premiums for motor third party liability insurances depend on the residence of policyholders and the country is divided into a few zones. Therefore the main reason this study started was to see if a division like that is out-of date by examining the expected compensation cost for each postal code. And if it is not out-of-date, then see if it is necessary to rearrange the postal codes in the zones.

The available data are examined over a period of 16 years. Data from one risk category were analyzed. Also, only results for property loss are given in this thesis.

The first aim of this study is to estimate the expected compensation cost the insurance company needs to pay due to motor liability insurances (category i) and ii), see before page) and drivers accident insurances (category iii)). To estimate the risk premiums, Bayesian statistics are used and statistical models are built for the expected claim frequencies and the expected total claim sizes which are used to compute the expected compensation cost. One of the main advantages of using a Bayesian statistics is that it is easy to obtain a certainity estimates for any function of the unknown parameters, such as the expected compensation cost. A Poisson model is used for the model for claim frequency and a gamma model is used for the total claim sizes. For more information about Bayesian statistics and these models see [1] and [9]. For parameter estimation Markov chain Monte Carlo (MCMC) is used, see [14] where the MCMC samplers have been implemented in Matlab. In this study the theory behind Gaussian Markov random fields (GMRF) is used to get information from neighbors, see more in [6]. In this case two neighborhood structures are constructed, one for time and other for postal codes. The neighborhood structures are described in Appendix A.

The second aim of this study is to include temporal and spatial factors and see if they improve the models. Having these two factors in the models allow for the evaluation of the time and postal codes effects. Also eight postal categories were selected depending on type of regions to improve the regional factor. These categories are similar to the zone separation the insurance company uses where the zones can be changed according to the results of the expected compensation cost.

No similar study has been conducted before here in Iceland. In Germany and Norway similar researches have been made, see [16] and [20] and this study is mainly based on those two papers. There, it is also preferred to have separate analysis of claim frequency and claim size. Similar to this study, a Poisson model is used for the expected claim frequency and a gamma model for the expected total claim sizes, and both models included spatial factors. Only spatial factor was used in those two papers but both spatial and temporal factors in this study. In both [16] and [20] claim sizes are examined per policyholder but in this study the data contains total claim size per postal code. [16] extends [20] but the before mentioned allows for dependencies between the number of claims and claim size.

1.1. Introduction to the thesis

The main references in this thesis are [16] and [20], that show similar researches abroad. Also [1], [9] and [7] are mainly used for the theory behind the work.

An outline of the thesis is given in the following. Chapter 2 introduces the theory behind the models used in this thesis. The basics of Bayesian inference and MCMC simulation methods are briefly summarized together with a description of the Poisson model and the gamma model. GMRF are introduced for the neighborhood structure, the Compound collective risk model is briefly discussed and finally model comparison is addressed. In Chapter 3, a presentation of the data used in the study is given. In Chapter 4 the models used in the methods for study are expressed. They are:

- Model for claim frequency
  - a Poisson model
- Model for total claim size
  - a gamma model
- Expected compensation
  - combination of Poisson and gamma models

Also prior distributions, the posterior distributions and conditional posterior distributions for MCMC are shown. The main results are given in Chapter 5, that is, results on expected claim frequency, expected total claim size and expected compensation.

2. Theory

In this chapter the theory used in this thesis is explained. It starts by giving a short introduction to the Poisson distribution and gamma distribution. Following there is an introduction to Gaussian Markov random field and the Compound collective risk model is described. Then Markov chain Monte Carlo (MCMC) simulation is briefly summarized and finally a short description of a model criteria called DIC is given.

2.1. The Poisson Model

Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed interval of time (or space) if these events occur with a known average rate and independently of the time since the last event. If the expected number of occurrences in this interval is λ, then the probability of exactly k occurrences is given by the formula for the Poisson probability mass function

$$p(k|\lambda) = \text{Poisson}(k|\lambda) = \frac{1}{k!}\lambda^k \exp(-\lambda), \qquad k \in \{0, 1, ..., \infty\},$$

where $\lambda > 0$. Here k denotes both the random variable k and a particular value of k, which one is returned will be clear from the context. The random variable k follows a Poisson distribution with parameter λ, denoted by $k \sim$Poisson(λ). In the Poisson model the mean and the variance of k are

$$E(k|\lambda) = \lambda$$

and

$$\text{Var}(k|\lambda) = \lambda.$$

[1] and [18] Figure 2.1 shows the probability mass function for the Poisson distribution for four different values of λ.

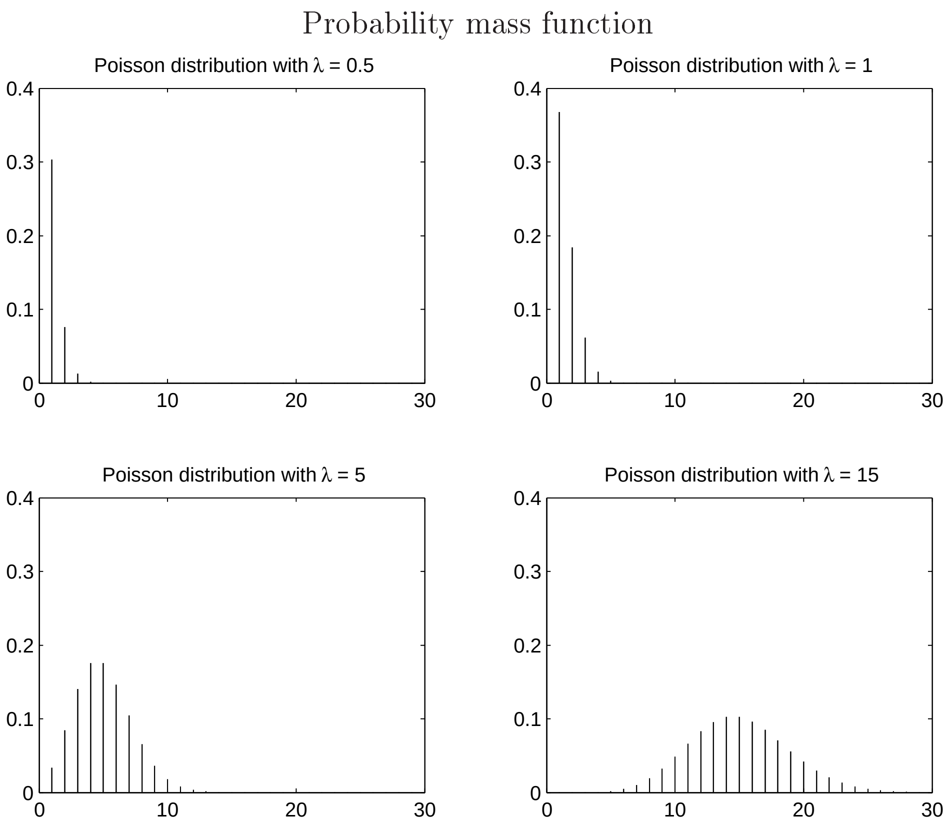

Figure 2.1: Poisson probability mass function with $\lambda = 0.5$, 1, 5 and 15. The horizontal axis is k.

2.2. The gamma Model

The gamma distribution is a two-parameter family of continuous probability distributions. It has a scale parameter β and a shape parameter α, $\alpha > 0$ and $\beta > 0$. [18] According to [1] the gamma distribution is the conjugate prior distribution for the inverse of the normal variance and for the mean parameter of the Poisson distribution. When $\alpha > 0$ the gamma integral is finite and the density function is finite. As $\alpha \to 0$ and $\beta \to 0$ a noninformative distribution is obtained in the limit.

$$\theta \sim \text{gamma}(\alpha, \beta)$$

denotes a random variable θ that follows a gamma distribution with parameters α and β.

$$p(\theta) = \text{gamma}(\theta|\alpha, \beta)$$

denotes the density function of the gamma distribution which is given by

$$p(\theta) = \frac{\beta^\alpha}{\gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}, \theta > 0$$

In the gamma model the mean and variance are

$$\text{E}(\theta) = \frac{\alpha}{\beta} \tag{2.1}$$

and

$$\text{Var}(\theta) = \frac{\alpha}{\beta^2} \tag{2.2}$$

Figure 2.2 shows the probability density function for four different values of α.

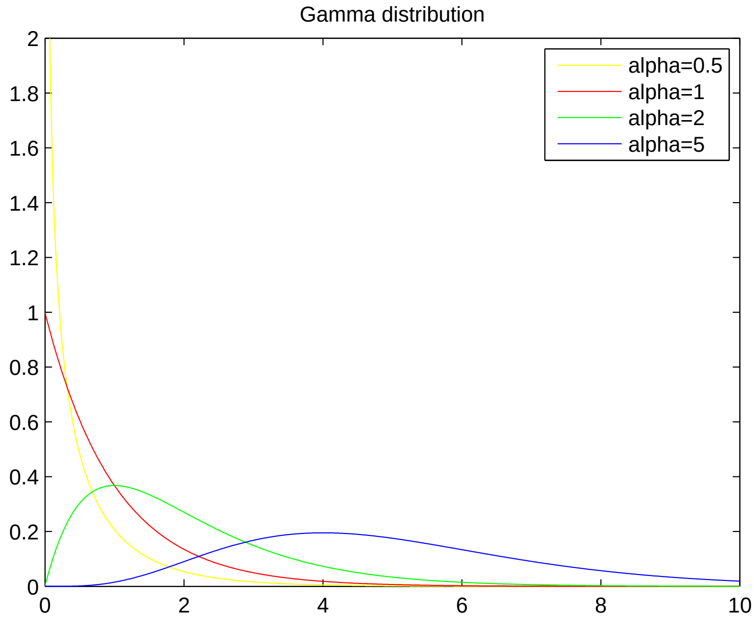

Figure 2.2: Gamma probability density function with α = 0.5, 1, 2 and 5.

2.3. GMRF

Gaussian Markov random fields (GMRF), also referred to as conditionally specified Gaussian models, can be useful for risk premium estimation because of the spatially interacting variables as mentioned in [20]. GMRF provides a convenient way to create a covariance matrices who describe the spatial and temporal correlation based on defining neighbors.
Let $Y = (Y_1, ..., Y_n)^T$ be an n vector of random variables which follows a Gaussian distribution such that

$$Y \sim N(\mu, (I - C)^{-1}M)$$

where $(I - C)$ is invertible and $(I - C)^{-1}M$ is symmetric and positive definite. C is an $n \times n$ matrix whose (i, j)-th element is $c_{ij}, c_{ij}\tau_j^2 = c_{ji}\tau_i^2, c_{ii} = 0$, M is a diagonal matrix, $M = \text{diag}(\tau_1^2, ..., \tau_n^2)$ and $\mu = (\mu_1, ..., \mu_n)^T$. Usually in models for areal data, neighboring units i and j are such that $c_{ij} \neq 0$ while for parts k and l that are far apart $c_{kl} = 0$, [6].

In this thesis the MRF are specified in the following way. Let H be a matrix describing the spatial structure between n units, and all of its elements are nonnegative. Let h_{ij} denote the (i, j)-th element of H. Then let M be a diagonal matrix such that

$$M_{i,j} = \frac{1}{\sum_{i=1}^{n} H_{i,j}}, \quad i = 1, ..., n$$

and in matrix form it will be

$$M = \begin{bmatrix} M_{11} & 0 & \cdots & 0 \\ 0 & M_{22} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & M_{nn} \end{bmatrix}$$

C is such that

$$C_{i,j} = \frac{H_{i,j}}{\sum_{j=1}^{n} H_{i,j}}$$

and

$$C = M \cdot H$$

The covariance matrix of Y is assumed to be

$$\Sigma = \tau^2 (I - \phi C)^{-1} M$$

where τ^2 is a variance parameter.

The ordered eigenvalues of C, $\lambda_{(j)}$, is such that $\lambda_{(1)} \leq \lambda_{(2)} \leq ... \leq \lambda_{(n)}$. Then the determinant of $(I - \phi C)$ can be written as

$$|I - \phi C| = \prod_{j=1}^{n} (1 - \phi \lambda_{(j)})$$

where ϕ is unknown scalar parameter that quantifies spatial dependence.

2.4. The compound collective risk model

In the basic insurance risk model, according to [8], the number of claims and the total claim produced by a portfolio in a given time period $t = 1, ..., T$ for some class i is denoted by (N_{it}, S_{it}) where

$$S_{it} = \sum_{k=1}^{N_{it}} W_{itk}, \quad \text{if } N_{it} > 0, \tag{2.3}$$

and zero otherwise, where W_{itk} is the amount of the kth claim at time t for some class i.

This model is due to [12] and has the following assumptions:

- The number of claims in the interval $(t-1, t)$, denoted by N_{it}, is a random variable

- The claim size $W_{itk}, k = 1, 2, ..., n_{it}$, is conditional on $N_{it} = n_{it}$. W_{itk} is positive independent and identically distributed random variables with finite mean $\mu_{it} = \mathrm{E}[W_{itk}]$ and variance $\sigma_{it}^2 = \mathrm{Var}(W_{ki}) < \infty$.

- The claim time occurs at random instants $t_{1i} \leq t_{2i} \leq \cdots$ 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 $\mathrm{E}[T_{ji}] = \lambda_i^{-1}$.

If it is assumed that the sequences T_j and W_j are independent from each other and the above conditions hold, then N_{it} is a homogeneous Poisson process with rate λ_{it}. Then assuming that W_{itk} is gamma$(\kappa_{it}, \theta_{it})$ distributed and that the inter-arrival times are exponentially distributed, the model is given by

$$\begin{array}{lcll} N_{it}|\lambda_{it}, \pi_{it} & \sim & \mathrm{Poisson}(\lambda_{it}\pi_{it}), & \lambda_i > 0, \\ X_{it}|n_{it}, \theta_{it} & \sim & \mathrm{gamma}(\kappa_{it}, \theta_{it}), & \theta_i > 0, \end{array}$$

where $\kappa_{it} = n_{it}\kappa_i, n_{it}$ is the observed number of claims at time t, for age class i, and π_{it} is the insured population at time t for age class i. This model is often referred to as the Poisson - gamma model for compound collective risk. [8]

2.5. Computation using Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) is an important technique used to simulate samples from a given density, for example the density of the posterior distribution of θ. The samples come in chains where each of the simulated values of θ depends on the preceding value. The basic principle is that when this chain has run sufficiently long enough it will represent the desired posterior distribution, $p(\theta|y)$. This distribution can be summarized by computing summary statistics from recorded values. The term Markov chain stands for a sequence of random variables $\theta^1, \theta^2, ...,$ for which, for any t, the distribution of θ^t given all previous θ's depends only on the most recent value, θ^{t-1}. The first step of MCMC simulation is to select a starting value θ^0 and then for each t, θ^t is drawn from a transition distribution, $T_t(\theta^t|\theta^{t-1})$ that depends on the last draw, θ^{t-1}, where $T_t(\theta^t|\theta^{t-1})$ must be constructed so that the Markov chain converges to the posterior distribution, $p(\theta|y)$, [9] and [1].

The two basic and most widely used algorithms used in MCMC are the Gibbs sampler and the Metropolis-Hastings algorithm. The Gibbs sampler was first introduced by [15] in 1984 and in 1990 [2] then showed how the method could be applied to a wide variety of Bayesian inference problems. The Metropolis-Hastings sampler was developed by [11] in 1953 and [19] in 1970. These algorithms are described in the following sections. For more informations see [17] and [1].

2.5.1. The Gibbs sampler

According to [9], the Gibbs sampler is a method to produce useful chain values. It requires specific knowledge about the relationship between the variables of interest. The basic idea is that if it is possible to express each of the parameters to be estimated as conditioned on all of the others, then by going through these conditional statements eventually the true joint distribution of interest is reached.

The Gibbs sampler, which is also called alternating conditional sampler, is defined in terms of subvectors of θ. [1] Assume the parameter vector θ has been devided into d subvectors, $\theta = (\theta_1, ..., \theta_d)$, then the objective is to produce a Markov chain that cycles through the subvectors of θ moving toward and then around this distribution. In each iteration t, a sample of each subvector is obtained by sampling from the distribution of the subvector conditioned on the latest value of the other subvector. Let

$$p(\theta_j | \theta_{-j}^{t-1}, y)$$

be the conditional distribution of θ_j, given the data and the other subvectors at their current value, denoted by θ_{-j}^{t-1} where

$$\theta_{-j}^{t-1} = (\theta_1^t, ..., \theta_{j-1}^t, \theta_{j+1}^{t-1}, ..., \theta_d^{t-1}).$$

The Gibbs sampler proceeds by selecting a starting value for θ (θ^0) and then by sampling from the d conditional distribution for each $t = 1...L$, where L is the number of iterations. The Gibbs sampler can therefore be presented as

$$\begin{gathered} \theta_1^t \; p(\theta_1^t | \theta_{-1}^{t-1}, y) \\ \theta_2^t \; p(\theta_2^t | \theta_{-2}^{t-1}, y) \\ \vdots \\ \theta_K^t \; p(\theta_K^t | \theta_{-K}^{t-1}, y). \end{gathered}$$

Once convergence is reached, all simulation values are from the target posterior distribution and a sufficient number should then be drawn so that all areas of the posterior are explored. During each iteration of the cycling through the θ vector, conditioning occurs on θ values that have already been sampled for that cycle - otherwise the θ values are taken from the last cycle.

2.5.2. Metropolis-Hastings Sampler

According to [9], the full set of conditional distributions for the Gibbs sampler are often quite easy to specify from the hierarchy of the model, since conditional relationship are directly in such statements. However, the Gibbs sampler obviously does not work when the complete conditionals for the θ parameters do not have an easily obtainable form. In these cases a chain can be produced for these parameters using the Metropolis-Hasings algorithm.

In [1] it is shown that the Metropolis-Hastings algorithm for the j-th subvector of parameters θ_j, in the t-th iteration is as follows:

i) Sample a proposal θ_j^* from a proposal distribution with density $J_{j,t}(\theta_j^*|\theta_j^{t-1})$.

ii) Calculate the ratio of the densities

$$r = \frac{p(\theta_j^*|\theta_{-j}^{t-1}, y)/J_{j,t}(\theta_j^*|\theta_j^{t-1})}{p(\theta_j^{t-1}|\theta_{-j}^{t-1}, y)/J_{j,t}(\theta_j^{t-1}|\theta_j^*)}$$

iii) Set $\theta_j^t = \theta_j^*$ with probability $\min(r, 1)$, otherwise set $\theta_j^t = \theta_j^{t-1}$.

Usually it is easier to work with r on the logarithmic scale in terms of numerical computation and for analytical results:

$$\log(r) = \log p(\theta_j^*|\theta_{-j}^{t-1}, y) - \log J_{j,t}(\theta_j^*|\theta_j^{t-1}) - \log p(\theta_j^{t-1}|\theta_{-j}^{t-1}, y) + \log J_{j,t}(\theta_j^{t-1}|\theta_j^*).$$

The Metropolis-Hastings step is an adaptation of a random walk, that uses an acceptance/rejection rule to converge to the specified target distribution. This acceptance rate is recommended to be 44% when θ_j is a scalar and 23% when θ_j is of higher dimension to ensure proper convergence and it is tuned by changing the variance of the proposal distribution.

2.6. Model comparison

For complex hierarchical models the computation of Bayes factors, according to [13], requires substantial efforts. Therefore it is helpful to consider model choice criteria which can easily be computed using the available MCMC output. Natural way to compare models is to use criterion based on trade-off between the fit of the data to the model and the corresponding complexity of the model. Proposed by [3], a Bayesian model comparison criterion based on this principle which will be described in next subsection.

2.6.1. Deviance Information Criterion (DIC)

The deviance information criterion, suggested by [3], for a probability model $p(y|\theta)$ with observed data $y = (y_1, \ldots, y_n)$ and unknown parameters θ is defined by

$$\mathrm{DIC} := \mathrm{E}[D(y, \theta|y)] + p_D. \tag{2.4}$$

which considers both model fit and model complexity where $p(y, \theta) = -2\log(p(y|\theta)$. The posterior mean of $D(y, \theta)$, $\mathrm{E}[D(y, \theta)|y]$ can be estimated with

$$\hat{D}_{avg}(y) = \frac{1}{L}\sum_{l=1}^{L} D(y, \theta)$$

and p_D is the effective number of parameters. The effective number of parameters, p_D measures the model complexity. p_D is estimated with

$$p_D := \hat{D}_{avg}(y) - D_{\hat{\theta}}(y). \tag{2.5}$$

The DIC criterion has been suggested as a criterion of model fit when the goal is to find a model that will be best for prediction when taking into account uncertainty due to sampling. As mentioned in [16], according to the DIC criterion the model with the smallest DIC is to be preferred. Both DIC and p_D can easily be computed by taking the posterior mean of the deviance $\hat{D}_{avg}(y)$ and the plug-in estimate of the deviance $D_{\hat{\theta}}(y)$ by using the available MCMC output. In this project the standardizing term $f(y)$ is zero.

3. Data

The data set which is analyzed in this thesis, comes from the icelandic insurance company Vátryggingafélag Íslands hf. (VÍS). The data sets contain the number of claims and total claim size, categorized by year, postal code and risk category.

3.1. Data description

The data set contains information about car insurance premium and compensation paid from the insurance company from 1993 to 2008. The number of policy days (which is changed to policy years when using the data), e_{it}, is known and not all policyholders were insured during the whole year. There are three main categories:

i) Property loss (i. munatjón)
- Only damage to properties, not people

ii) Bodily injury (i. líkamstjón)
- injury to people other than driver and owner of car

iii) Drivers accident insurance (i. slysatrygging ökumanns og eiganda)
- Only injury to owner and driver of the car

The data sets contain the following variables:

- Year when claim occurs
- Risk category
- Region (postal code)
- Premiums for liability insurance
- Premium for driver accident insurance

- Number of policy days
- And for three categories i), ii) and iii)
    - Estimated total compensation
    - Number of claims
    - Estimated unpaid compensation

Iceland is divided into 151 regions, including post office box. For convenience post office box were combined with appropriate postal codes, reducing the number of regions used to 130. For each claim the year claim occurs, t, and region where policyholder is residing, i, is known. The original data set contains about 35.500 observations, but in this thesis only one risk category is analyzed. The data set for this category contains about 2230 observations when regions are 151 but about 2000 when regions are 130. All data have been brought to present worth but they have also been scaled for reasons of confidentiality.
In the model for claim frequency only observation with non-zero policy years are taken into account. Of course, if the number of policy years is greater than zero then account of zero for particular postal code and particular year is taken into account. In case of the model for the claim size, observations with non-zero policy years are taken into account for all occurred claims. There is a quite large amount of observations with no claim, especially in the category for divers accident insurance, see Tables 3.1, 3.2 and 3.3. They also show the maximum number of observed claims, which is most in the category for property loss.

Table 3.1: Summary of the observed claim frequencies in the data of property loss.

Property loss	
Number of claims	Percentage of observation
0	43.72
1	17.38
2	9.01
3	4.51
4	3.00
⋮	⋮
63	0.05
64	0.05

In Figure 3.1 a histogram of the observed positive average claim sizes is given and in Table 3.4 is given the mean of total claim sizes per number of claims. The figure shows the average claim size is just below 100000. From the mentioned table the total

Table 3.2: Summary of the observed claim frequencies in the data of bodily injury.

Bodily injury	
Number of claims	Percentage of observation
0	77.82
1	9.71
2	5.11
3	2.65
4	1.65
$\vdots$	$\vdots$
13	0
14	0.05

Table 3.3: Summary of the observed claim frequencies in the data of drivers accident insurance.

Drivers accident insurance	
Number of claims	Percentage of observation
0	83.93
1	10.82
2	3.15
3	1.45
4	0.50
5	0.05
6	0.05
7	0.05

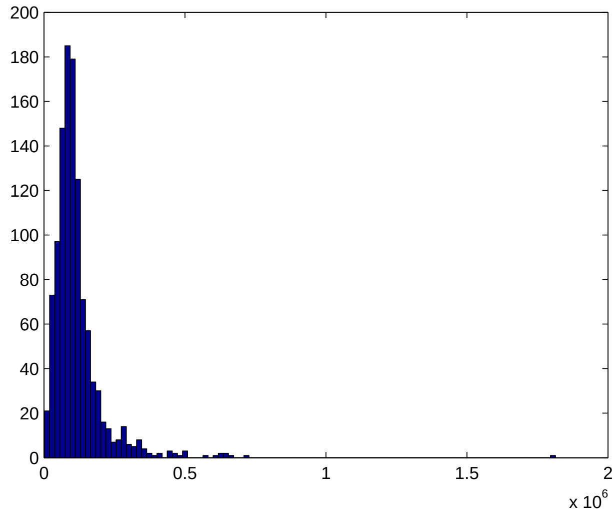

Figure 3.1: Histogram of the observed positive total claim sizes per number of claims.

claim size per number of claims is 93044. The largest average claim size observation per number of claim is 1451240 which is about 1.4% of the sum of all total claim sizes per number of claims. Based on a simple statistical test no relationship was found between number of claims and total claim sizes. Figure 3.1 and Table 3.4 are therefore describing for the data.

Table 3.4: The mean of total claim sizes per number of claims taken over all observations and over observations with $N = k, k = 1, 2, 3, 4, 5, 6, 7, 8, 63, 64$.

Number of observations	Mean
All	93044
N = 1	94220
N = 2	93970
N = 3	90564
N = 4	96060
N = 5	105416
N = 6	95580
N = 7	96583
N = 8	89150
⋮	⋮
N = 63	76020
N = 64	107956

3.1.1. Neighborhood matrix

In Chapter 2.3 a model describing correlation between variables based on their neighborhood structure was given. In this subsection, the matrix that describes the neighborhood structure is defined. If postal codes are neighbors, 1 connects them, otherwise there is 0. Postal code has to have at least one neighbor and is never neighbor to it self. The neighborhood matrixes can be seen in Appendix A.
There are two kind of neighborhood structures, temporal and regional. In the neighborhood structure for time, years before and after a certain year are neighbors so each year has 2 neighbors except the first and last that only have one. The neighborhood structure for regions is defined such that regions are neighbors if they:

- share borders
- share borders with neighbors when there is strong connection
  - strong connection within the capital area
  - strong connection if postal codes are within the same commune

4. Models

Under the Poisson-gamma model for compound collective risk, see Section 2.4 the number of claims follow a Poisson distribution and are independent of the claim sizes which follow a gamma distribution. In this chapter these two models are described. The main result is expected compensation which is simply the expected claim frequency times the expected total claim size. Following there is an introduction to the Bayesian inference where prior- and posterior distributions and MCMC algorithms are described.

4.1. Model for claim frequency

For claim frequency, a Poisson model with spatial and temporal effects is chosen. This model is easy to use for insurance data, e.g. based on [20] and [16]. The structure of the data requires spatial and temporal effects which are easy to include to the proposed model. Considering only observations with non-zero policy years, altogether 1974 observations are obtained. The index t denotes the year claim occurs and the index i denotes the region where policyholder is residing. N_{it} is the number of claims in region i at year t. The number of policy years for policyholders in region i and at year t is denoted by e_{it}. The average number of claims per a single policy year in region i at year t is denoted by θ_{it}. The proposed model for number of claims is given by

$$N_{it} \mid \theta_{it} \sim \text{Poisson}(e_{it} * \theta_{it}) \tag{4.1}$$

or with $\theta^*_{it} = \log \theta_{it}$

$$N_{it} \mid \theta^*_{it} \sim \text{Poisson}(e_{it} * \exp(\theta^*_{it}))$$

where

$$\theta^*_{it} = x_i^T \beta + a_{1,t} + a_{2,i} + \varepsilon_{it},$$

$a_{1,t}$ and $a_{2,i}$ are temporal- and spatial factors respectively. The temporal factor is used to see if time influences the results and the spatial factor is to see if residence of policyholders influencees the results. x_i is a vector of covariates, β is a vector of parameters and ε_{it} is an error term which is independent over years and postal codes,

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma^2_{\varepsilon,it})$$

where

$$\sigma^2_{\varepsilon,it} = \log\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{4}{re_{it}}}\right)$$

with r denoting the average frequency per policy year. The calculations for $\sigma^2_{\varepsilon,it}$, which changes for each policy year, can be seen in Appendix B.

To estimate the number of claims, θ^*_{it}, eight potential postal categories were selected and used as covariates to help explain the spatial patterns of the claims. The vector of unknown regression parameters is $\beta = (\beta_1, ..., \beta_8)'$. The vector of covariates for the i-th observation is given by

$$x_i = (x_{1,i}, ..., x_{8,i})'$$

and includes only 0 and 1. The main role of this matrix is to ensure that each observation is connected to the correct postal category. When deciding the postal categories, the postal codes were grouped in several ways where the groups depended on sizes of urban regions, closeness to highway, etc. The model was tested under different grouping of the postal codes. The best result was found with help of DIC. The best result consisted of eight categories which can be seen in Table 4.1.

Table 4.1: Regression covariates.

Regression Covariates		
Covariates (x)	Coefficients (β)	Postal categories
x_1	β_1	Reykjavík
x_2	β_2	Reykjavík Urban Region
x_3	β_3	Large Urban Region
x_4	β_4	Small Urban Region
x_5	β_5	Rural near highway
x_6	β_6	Rural area
x_7	β_7	Banks and government
x_8	β_8	Other

4.1.1. Residual analysis

Evaluating the fit of the model to the data is an important step when building a model. Graphical residual analysis is the primary statistical tool to evaluate the fit of a model. Residuals are the difference between the measured output from the validation data set and the predicted model output. Residuals therefore represent the portion of the validation data not explained by the model. [10] When random errors in regression models are not normally distributed it can be useful to use alternative formula to the usual residual formula, proposed by [4]. It is based on the idea to produce quantities that are close to being normal distributed, and the formula for the residuals depends on the assumed distribution. For Poisson distribution, formula for the residual is given by

$$e = \frac{\frac{3}{2}(y^{2/3} - \psi^{2/3})}{\psi^{1/6}} \tag{4.2}$$

where y follows a Poisson distribution with mean ψ. Modify for the overdispersed Poisson model so that the expected value and the variance of N_{it} are conditioned on β, a_1 and a_2 (integrated over e_{it}). The mean is

$$\mathrm{E}(N_{it}|\beta, a_1, a_2) = e_{it}\exp(X_i^T\beta + a_{1,t} + a_{2,i} + 0.5\sigma^2_{\varepsilon,it}) = \psi_{it}$$

and the variance is

$$\mathrm{Var}(N_{it}|\beta, a_1, a_2) = \psi^2_{it}(\exp(\sigma^2_{\varepsilon,it}) - 1) + \psi^2_{it}.$$

To fit Equation (4.2) to the model for frequency of claims, according to Taylor expansion the equation becomes

$$e = \frac{y^{2/3} - \psi^{2/3}}{\sqrt{\sigma^2_Y \frac{4}{9}\psi^{-2/3}}} \tag{4.3}$$

and by using Equations (4.2) and (4.3) together with using the transformation of $N_{it}^{2/3}$ and Cramérs Theorem, [5] the residual formula for the claim frequency model, where y is denoted by N_{it} is

$$e = \frac{N_{it}^{2/3} - \exp(\psi_{it}^{2/3})}{\sqrt{\psi^2_{it}(\exp(\sigma^2_{\varepsilon,it}) - 1) + \psi_{it}}} * \frac{1}{\frac{2}{3}(\psi_{it})^{-1/3}}. \tag{4.4}$$

4.2. Model for claim size

The proposed model for claim size is a gamma model with temporal and spatial effects and is similar to the model presented in [20] and [16]. The experience shows this kind of model is easy to use for similar projects. The structure of the data requires temporal and spatial effects. These effects are easy to include to the proposed

model. In this model only observations with positive number of claims and non-zero policy years are taken into account, altogether 1108 observations are obtained. S_{it} is the total size of claims in region i and year t. The average size of each claim is denoted by μ_{it} and σ_{it} denotes the standard deviation of each claim.
The original idea for this model is similar to the model for claim frequency. The log-mean parameter μ^* is similar to θ^* and the factors d_1 and d_2 are similar to a_1 and a_2, i.e. temporal and spatial effects respectively. The log-mean is modeled as $X_i^T\eta + d_{1,t} + d_{2,i}$ but in the final analysis d_2 was cut out from the model. Nonetheless d_2 will be included in this chapter, but will be discussed further in Chapter 5.

The covariate vector is the same as for the model for claim frequency, $x_i = (x_{i1}, ..., x_{i8})'$, only containing 0 and 1. The vector of unknown regression parameters is $\eta = (\eta_1, ..., \eta_8)'$.

$$S_{it} \mid N_{it} \sim \text{gamma}(N_{it}\mu_{it}, N_{it}\sigma_{it}^2) \tag{4.5}$$

where the gamma distribution is parameterized by its mean and variance, and

$$\mu_{it} = \exp(x_i^T\eta + d_{1,t} + d_{2,i})$$

$$\sigma_{it}^2 = \frac{\mu_{it}^2}{\alpha} \quad \text{where } \alpha \text{ is an unknown parameter}$$

$$S_{it} = \sum_{k=1}^{N_{it}} W_{itk}, \qquad W_{k,it} \sim \text{gamma}(\mu_{it}, \sigma_{it}^2)$$

where W_{itk} is the cost for individual claim, which is unknown.

Like for claim frequency, model validation with graphical analyzation of the residuals is made for the claim sizes. Also proposed by [4], a model for residuals in case of gamma distribution with mean μ

$$e = \frac{3(y^{1/3} - \mu^{1/3})}{2\mu^{1/3}}. \tag{4.6}$$

The formula in 4.6 needs to be corrected by including the square root of α. That is by letting

$$e = \frac{3\sqrt{\alpha}(y^{1/3} - \mu^{1/3})}{2\mu^{1/3}}, \tag{4.7}$$

the variance of e becomes approximately 1. In case of the model for S_{it}

$$e = \frac{3\sqrt{\alpha N_{it}}(S_{it}^{1/3} - N_{it}^{1/3}\mu_{it}^{1/3})}{2N_{it}^{1/3}\mu_{it}^{1/3}} \tag{4.8}$$

with α_{it} and β_{it} are such that

$$\alpha_{it} = \alpha \cdot N_{it},$$
$$\beta_{it} = \alpha \cdot \exp(-\mu)$$

where α_{it} and β_{it} are the usual parameters of the gamma distribution and

$$\mu_{it} = \frac{\alpha_{it}}{\beta_{it}}.$$

4.3. Expected compensation

The main aim of the research in this thesis is to find out the expected total compensation cost the insurance company has to pay. This cost will depend on models for claim frequency and claim size, and is examined for each postal code where policyholder is residing. The expected value is the expected total cost (S_{it}) per number of policy years (e_{it}) and the final result is a combination of the two main models mentioned earlier in this chapter. Despite of this combination, the uncertainty does not increase since each part is calculated separately.

$$\begin{aligned}
E\left(\frac{S_{it}}{e_{it}}\right) &= \tfrac{1}{e_{it}}E(S_{it}) \\
&= \tfrac{1}{e_{it}}E\left(\textstyle\sum_{k=1}^{N_{it}} W_{k,it}\right) \\
&= \tfrac{1}{e_{it}}E(E(\textstyle\sum_{k=1}^{N_{it}} W_{k,it}|N_{it})) \\
&= \tfrac{1}{e_{it}}E(\textstyle\sum_{k=1}^{N_{it}} E(W_{k,it})|N_{it}) \\
&= \tfrac{1}{e_{it}}E(N_{it}E(W_{1,it})) \\
&= \tfrac{1}{e_{it}}E(N_{it})E(W_{1,it}) \\
&= \tfrac{1}{e_{it}}E(E(N_{it}|\theta_{it}^*))\mu_{it} \\
&= \tfrac{1}{e_{it}}E(e_{it}\theta_{it}^*)\exp(x_i^T\eta + d_{1,t} + d_{2,i}) \\
&= \exp(x_i^T\beta + a_{1,t} + a_{2,i} + \tfrac{1}{2}\sigma_\varepsilon^2)\exp(x_i^T\eta + d_{1,t} + d_{2,i})
\end{aligned} \tag{4.9}$$

where $\exp(x_i^T\beta+a_{1,t}+a_{2,i}+\frac{1}{2}\sigma_\varepsilon^2)$ is the Poisson part of the model and $\exp(x_i^T\eta+d_1,t+d_{2,i})$ is the gamma part of the model, see Equations (4.2) and (4.6), respectively.

4.4. Bayesian inference

4.4.1. Prior distributions

An important part of Bayesian inference is the selection of appropriate prior distributions for the unknown parameters. For the model for claim frequency the prior distributions for the parameters mentioned in Section 4.1 are assumed to be as follows:

$$\begin{aligned}
\beta &\sim N(\mu_\beta, \sigma_\beta^2 I) \\
a_1 &\sim N(\bar{0}, \tau_1^2(I-\phi_1 C_1)^{-1}M_1) \\
a_2 &\sim N(\bar{0}, \tau_1^2(I-\phi_2 C_2)^{-2}M_2) \\
\varepsilon_{it} &\sim N(\bar{0}, \sigma_{\varepsilon,it}^2 I)
\end{aligned} \tag{4.10}$$

where $N(\cdot,\cdot)$ indicates the normal distribution. μ_β and σ_β are selection based pre-analysis of data. β is estimated with Maximum likelihood estimation in a model with $\theta^* = x_i\beta$. μ_β is the mean of $\hat{\beta}$ and σ_β^2 is C times the variance of $\hat{\beta}$, where C is a multiplication factor for the variance decided with experiments on the model with default value equal to 4 in the model for claim frequency and equal to 0.5 for the model for claim size. This is done to get the estimation of β stable.

The hyperparameters are $\sigma_\varepsilon^2, \tau_1^2, \tau_2^2, \phi_1$ and ϕ_2. A non-informative prior distribution - scaled inverse-χ^2 is chosen for $\sigma_{\varepsilon,it}^2, \tau_1^2$ and τ_2^2. For ϕ_1 and ϕ_2 beta distribution is chosen to ensure strong correlation.

$$\begin{aligned}
\sigma_\varepsilon^2 &\sim \text{Inv} - \chi^2(\nu_\varepsilon, S_\varepsilon^2) & \nu_\varepsilon &= 10^{-6}, & S_\varepsilon^2 &= 1 \\
\tau_1^2 &\sim \text{Inv} - \chi^2(\nu_{\tau 1}, S_{\tau 1}^2) & \nu_{\tau 1} &= 10^{-6}, & S_{\tau 1}^2 &= 1 \\
\tau_2^2 &\sim \text{Inv} - \chi^2(\nu_{\tau 2}, S_{\tau 2}^2) & \nu_{\tau 2} &= 10^{-6}, & S_{\tau 2}^2 &= 1 \\
\phi_1 &\sim beta(\alpha_{\phi 1}, \beta_{\phi 1}) & \alpha_{\phi 1} &= 100, & \beta_{\phi 1} &= 0.5 \\
\phi_2 &\sim beta(\alpha_{\phi 2}, \beta_{\phi 2}) & \alpha_{\phi 2} &= 100, & \beta_{\phi 2} &= 0.5
\end{aligned} \tag{4.11}$$

Similar for claim size, the prior distributions are as follows:

$$\begin{aligned}
\eta &\sim N(\mu_\eta, \sigma_\eta^2 I) \\
d_1 &\sim N(\bar{0}, \kappa_1^2 (I - \zeta_1 C_1)^{-1} M_1) \\
d_2 &\sim N(\bar{0}, \kappa_2^2 (I - \zeta_2 C_2)^{-1} M_2) \\
\alpha &\sim \text{gamma}(\alpha_0, \beta_0)
\end{aligned} \tag{4.12}$$

where $N(\cdot, \cdot)$ indicates the normal distribution. The parameters in the gamma model are estimated similar to the parameters in the model for claim frequency. μ_η and σ_η are selection based pre-analysis of data. η is estimated with Maximum likelihood estimation in a model $x_i\eta$. μ_η is the mean of $\hat{\eta}$. For α gamma distribution is chosen with parameters $\alpha_0 = 1$ and $\beta_0 = 1$. The hyperparameters are $\sigma_r^2, \kappa_1^2, \kappa_2^2, \zeta_1$ and ζ_2. A non-informative prior distribution - scaled inverse-χ^2 is chosen for σ_r^2, κ_1^2 and κ_2^2. For ζ_1 and ζ_2 beta distribution is chosen to ensure strong correlation.

$$\begin{aligned}
\sigma_r^2 &\sim Inv - \chi^2(\nu_r, S_r^2) & \nu_r &= 10^{-6}, & S_r^2 &= 1 \\
\kappa_1^2 &\sim Inv - \chi^2(\nu_{\kappa 1}, S_{\kappa 1}^2) & \nu_{\kappa 1} &= 10^{-6}, & S_{\kappa 1}^2 &= 1 \\
\kappa_2^2 &\sim Inv - \chi^2(\nu_{\kappa 2}, S_{\kappa 2}^2) & \nu_{\kappa 2} &= 10^{-6}, & S_{\kappa 2}^2 &= 1 \\
\zeta_1 &\sim beta(\alpha_{\zeta 1}, \beta_{\zeta 1}) & \alpha_{\zeta 1} &= 100, & \beta_{\zeta 1} &= 0.5 \\
\zeta_2 &\sim beta(\alpha_{\zeta 2}, \beta_{\zeta 2}) & \alpha_{\zeta 2} &= 100, & \beta_{\zeta 2} &= 0.5
\end{aligned} \tag{4.13}$$

4.4.2. Posterior distributions

The posterior distributions of the parameters, given the observed claim numbers and claim sizes, describe the statistical uncertainty and is the tool for inference in a Bayesian analysis. Analytically, the posterior density is the product of the prior density and the likelihood. The posterior distribution for the model of claim frequency is

$$
\begin{aligned}
p(\theta^*_{it}, a_1, a_2, \beta, \sigma^2_{\varepsilon,it}, \tau_1^2, \tau_2^2, \phi_1, \phi_2 | N_{it}) \quad \propto \quad & p(N_{it}|\theta^*_{it}) \\
& \times p(\theta^*_{it}|a_1, a_1, \beta, \sigma^2_{\varepsilon,it}) \\
& \times p(a_1|\tau_1^2, \phi_1) \times p(a_2|\tau_2^2, \phi_2) \\
& \times p(\beta)p(\sigma^2_{\varepsilon,it})p(\tau_1^2)p(\tau_2^2)p(\phi_1)p(\phi_2) \\
\propto \quad & \textstyle\prod_{i=1}^{J} \prod_{t=1}^{T} \text{Poisson}(N_{it}|e_{it}\theta^*_{it}) \\
& \times \textstyle\prod_{i=1}^{J} \prod_{t=1}^{T} \text{N}(\theta^*|x_i^T\beta + a_{1,t} + a_{2,i}, \sigma^2_{\varepsilon,it}) \\
& \times \text{N}(a_1|\underline{0}, \tau_1^2(I - \phi_1 C_1)^{-1} M_1) \\
& \times \text{N}(a_2|\underline{0}, \tau_2^2(I - \phi_2 C_2)^{-1} M_2) \\
& \times \text{N}(\beta|\mu_\beta, \sigma_\beta^2 I)\text{Inv-}\chi^2(\sigma_\varepsilon^2|\nu_\varepsilon, S_\varepsilon^2) \\
& \text{Inv-}\chi^2(\tau_1^2|\nu_{\tau 1}, S_{\tau 1}^2)\text{Inv-}\chi^2(\tau_2^2|\nu_{\tau 2}, S_{\tau 2}^2) \\
& \times \text{beta}(\phi_1|\alpha_{\phi 1}, \beta_{\phi 1})\text{beta}(\phi_2|\alpha_{\phi 2}, \beta_{\phi 2})
\end{aligned}
\tag{4.14}
$$

The posterior distribution for the claim sizes is

$$\begin{aligned}
p(d_1, d_2, \eta, \kappa_1^2, \kappa_2^2, \zeta_1, \zeta_2, \alpha | S_{it}) &\propto p(S_{it}|\eta, d_1, d_2, \alpha, N_{it}) \\
&\quad \times p(d_1|\kappa_1^2, \zeta_1) \times p(d_2|\kappa_2^2, \zeta_2) \\
&\quad \times p(\eta)p(\alpha)p(\zeta_1)p(\zeta_2) \\
&\propto \textstyle\prod_{i=1}^{J} \prod_{t=1}^{T} \text{gamma}\left[S_{it}|N_{it}\exp(\lambda_{it}), \frac{N_{it}}{\alpha}\exp(2\lambda_{it})\right] \\
&\quad \times \text{N}(d_1|\underline{0}, \kappa_1^2(I - \zeta_1 C_1)^{-1} M_1) \\
&\quad \times \text{N}(d_2|\underline{0}, \kappa_2^2(I - \zeta_2 C_2)^{-1} M_2) \\
&\quad \times \text{N}(\eta|\mu_\eta, \sigma_\eta^2 I) \times \text{gamma}(\alpha_0, \beta_0) \\
&\quad \times \text{Inv-}\chi^2(\kappa_1^2|\nu_{\kappa 1}, S_{\kappa 1}^2)\text{Inv-}\chi^2(\kappa_2^2|\nu_{\kappa 2}, S_{\kappa 2}^2) \\
&\quad \times \text{beta}(\zeta_1|\alpha_{\zeta 1}, \beta_{\zeta 1})\text{beta}(\zeta_2|\alpha_{\zeta 2}, \beta_{\zeta 2})
\end{aligned} \tag{4.15}$$

When finding the posterior distribution for the model for claim sizes first it is good to parameterize the model with the mean and variance. In Chapter 2 are the basic equations for the gamma model expressed, see Equations (2.1), (2.1) and (2.2). The usual parametrization is as follows:

$$\frac{\alpha}{\beta} = \mu,$$
$$\frac{\alpha}{\beta^2} = \sigma^2 = \frac{1}{\alpha}\frac{\alpha^2}{\beta^2} = \frac{1}{\alpha}\mu^2.$$

This gives

$$\alpha = \frac{\mu^2}{\sigma^2}$$
and
$$\beta = \frac{\alpha}{\mu} = \frac{1}{\mu}\frac{\mu^2}{\sigma^2} = \frac{\mu}{\sigma^2}. \tag{4.16}$$

Then the parametrization for this model is

$$\begin{aligned}
\lambda_{it} &= x_i^T \eta + d_{1,t} + d_{2,i}, \\
\mu_{it} &= N_{it} \exp(\lambda_{it}), \\
\sigma_{it}^2 &= \tfrac{1}{\alpha} N_{it} \exp(2\lambda_{it}), \\
\alpha_{it} &= \frac{N_{it}^2 \exp(2\lambda_{it})}{\frac{1}{\alpha} N_{it} \exp(2\lambda_{it})} = \alpha N_{it}, \\
\beta_{it} &= \frac{N_{it} \exp(\lambda_{it})}{\frac{1}{\alpha} N_{it} \exp(2\lambda_{it})} = \alpha \exp(-\lambda_{it}).
\end{aligned} \tag{4.17}$$

This parametrization is used in the calculations for the conditional distributions of the claim size model in Subsection 4.4.3.

4.4.3. MCMC

As mentioned in Section 2.5, Gibbs sampler is one kind of MCMC simulations. Gibbs sampler is used to simulate samples from the posterior distributions. It is sufficient to have 4 chains. Starting values are determined with experiments and then the MCMC algorithms are run for 15000 iterations. A burn-in of 4500-5000 iterations is found to be sufficient after experiments on the MCMC trace plots.

Sampling from the posterior distribution by using the Gibbs sampler gives the conditional distribution for all parameters. In Appendix C, conditional distributions for all parameters are expressed.

5. Results

In this chapter the main results of an analysis based on the derived methods are given. First, results on the expected claim frequency per policy year, expected total claim size per policy year and the expected compensation cost per policy year, both for the postal categories and for each postal code. Following, there are results from the claim frequency model and results from the claim size model. Finally, results on the expected compensation cost, which is based on these two models, is introduced. In Chapter 3, three categories for mandatory liability insurances for vehicles are mentioned. Here, only results for one category, property loss, will be listed out, both for claim frequency and claim size. Also, there are three postal codes that have no neighbors. According to the postal categories they belong to banks and government (postal codes 150 and 155) and other (postal code 999). These postal codes are not included in figures in this chapter and grouped separately in the postal categories.

5.1. Main results

In Chapter 4, eight different postal categories were introduced. Table 5.1 shows for these postal categories their expected claim frequency per number of policy years, their expected claim size per number of policy years and their expected compensation per number of policy years. This table shows the number of claims is highest in Reykjavík, urban regions around Reykjavík and Large urban regions but lowest in the rural areas near highway and the same is for expected compensation. In Appendix D, Tables D.1-D.3 are similar to Table 5.1, but these tables also give upper and lower bounds for 95% posterior interval. For example for claim frequency and category 1 the lower 95% interval is 0.061 which is larger than e.g. category 4 which has the upper 95% interval as 0.056 and therefore shows there is significant difference between these categories. It is interesting that the cost is higher in rural area than rural near highway. These two postal categories were examined further in terms of what postal codes were included. Based on this examination there was no reason to make changes. Claim size on the other hand is very similar between postal categories.

Table 5.1: Expected claim frequency, total claim size and compensation per policy year for the eight postal categories.

Numeric code	Region	Claim frequency	Claim size	Expected compensation
1	Reykjavík	0.0743	82290	6113
2	Reykjavík urban region	0.0663	82001	5440
3	Large urban region	0.0644	86078	5539
4	Small urban region	0.0513	88740	4551
5	Rural near highway	0.0460	86080	3956
6	Rural area	0.0505	86729	4377
7	Banks and government	0.1101	93742	10321
8	Other	0.0751	88912	6674

Table 5.2 is similar to Table 5.1. However, instead of postal categories all postal codes are expressed. The results are in line with results in Table 5.1, with the highest frequency and expected compensation in the capital area and lowest in the countryside. Size of claims are very similar between postal categories and are the same for postal codes within the same postal category. In Appendix D, Tables D.4 - D.6 are similar to Table 5.2 but these tables also give upper and lower bounds for 95% posterior intervals. For claim frequency the average lower bound multiplication factor is 0.6583 (about 34% is subtracted from the posterior mean) and the upper bound multiplication factor is 1.4593 (about 46% is added to the posterior mean). For claim sizes the lower bound multiplication factor is 0.8493 (about 15% is subtracted from the posterior mean) and the upper bound multiplication factor is 1.1899 (about 19% is added to the posterior mean). For expected value the lower bound multiplication factor is 0.6341 (about 37% is subtracted from the posterior mean) and upper bound multiplication factor is 1.5083 (about 51% is added to the posterior mean).

For better explanation, the expected compensation cost is shown graphically for all postal codes in Figure 5.1. There is a good overview for the numbers in the second last column in Table 5.2. Since the expected claim frequency reflects the expected compensation, and the expected total claim sizes are similar for all postal codes, it was decided only to show the expected compensation graphically, like in Figure 5.1. As can be seen, a few postal codes outside of the capital area have a large amount of expected compensation, even though it is a small region (for example postal codes 233, 345 and 611). This is explained by unusual amount of occurred claims during some years when compared to other years for that postal code (where the amount of policyholders is still the same) and the frequency of claims in nearby postal codes.

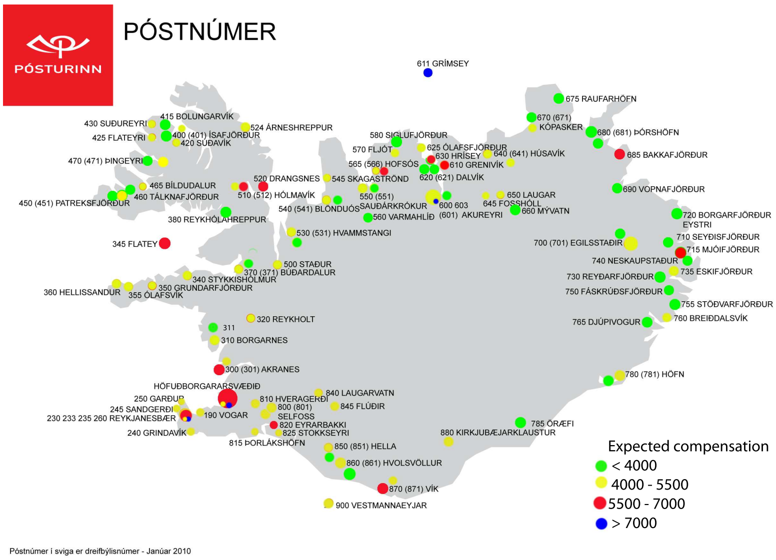

Figure 5.1: Expected compensation cost for each postal code. Green dot present expected compensation less than 4000, yellow dot present postal codes where expected compensation are between 4000 and 5500. Red dot present expected compensation between 5500 and 7000 and finally the blue one present expected compensation higher than 7000.

Table 5.2: Table with posterior mean for expected claim frequency per policy year (Poisson), expected total claim size per policy year (gamma) and expected compensation cost per policy year.

Region	Postal code	Claim frequency	Claim size	Expected compensation	Numeric code
Reykjavík	101	0.0761	82291	6265	1
Reykjavík	103	0.0713	82291	5871	1
Reykjavík	104	0.0771	82291	6345	1
Reykjavík	105	0.0696	82291	5725	1
Reykjavík	107	0.0681	82291	5606	1
Reykjavík	108	0.0702	82291	5776	1
Reykjavík	109	0.0711	82291	5848	1
Reykjavík	110	0.0718	82291	5912	1
Reykjavík	111	0.0821	82291	6753	1
Reykjavík	112	0.0622	82291	5117	1
Reykjavík	113	0.0678	82291	5583	1
Reykjavík	116	0.1040	82291	8558	1
Reykjavík	150	0.1146	93742	10743	7
Reykjavík	155	0.1056	93742	9898	7
Seltjarnarnes	170	0.0753	82001	6177	2

Continued on next page

Table 5.2 – continued from previous page

Region	Postal code	Claim frequency	Claim size	Expected compensation	Numeric code
Vogar	190	0.0593	88740	5261	4
Kópavogur	200	0.0652	82001	5344	2
Kópavogur	201	0.0607	82001	4976	2
Kópavogur	203	0.0632	82001	5185	2
Garðabær	210	0.0624	82001	5120	2
Hafnarfjörður	220	0.0716	82001	5870	2
Hafnarfjörður	221	0.0619	82001	5075	2
Álftanes	225	0.0716	82001	5868	2
Reykjanesbær	230	0.0717	86079	6174	3
Reykjanesbær	233	0.1185	86079	10204	3
Reykjanesbær	235	0.0545	86079	4689	3
Grindavík	240	0.0569	88740	5049	4
Sandgerði	245	0.0548	88740	4865	4
Garður	250	0.0568	88740	5040	4
Reykjanesbær	260	0.0625	86079	5376	3
Mosfellsbær	270	0.0652	82001	5346	2
Akranes	300	0.0641	86079	5521	3
Akranes	301	0.0556	86080	4789	5
Borgarnes	310	0.0489	86079	4212	3
Borgarnes	311	0.0454	86080	3912	5
Reykholt	320	0.0544	86729	4717	6
Stykkishólmur	340	0.0508	88740	4509	4
Flatey á Breiðarfirði	345	0.0728	86080	6261	6
Grundarfjörður	350	0.0459	88740	4072	4
Ólafsvík	355	0.0553	88740	4903	4
Snæfellsbær	356	0.0760	88740	6747	4
Hellissandur	360	0.0518	88740	4594	4
Búðardalur	370	0.0459	88740	4071	4
Búðardalur	371	0.0395	86729	3430	6
Reykhólahreppur	380	0.0424	86729	3675	6
Ísafjörður	400	0.0423	88740	3758	4
Ísafjörður	401	0.0527	86729	4567	6
Hnífsdalur	410	0.0442	86729	3834	6
Bolungarvík	415	0.0450	88740	3993	4
Súðavík	420	0.0539	88740	4780	4
Flateyri	425	0.0496	88740	4400	4
Suðureyri	430	0.0492	88740	4364	4
Patreksfjörður	450	0.0422	88740	3741	4

Continued on next page

Table 5.2 – continued from previous page

Region	Postal code	Claim frequency	Claim size	Expected compensation	Numeric code
Patreksfjörður	451	0.0476	86729	4126	6
Tálknafjörður	460	0.0364	88740	3230	4
Bíldudalur	465	0.0545	88740	4839	4
Þingeyri	470	0.0416	88740	3688	4
Þingeyri	471	0.0488	86729	4232	6
Staður	500	0.0499	86080	4292	5
Hólmavík	510	0.0478	88740	4238	4
Hólmavík	512	0.0708	86729	6139	6
Drangsnes	520	0.0666	88740	5911	4
	522	0.0302	86729	2618	6
	523	0.0636	86729	5519	6
Árneshreppur	524	0.0842	86729	7299	6
Hvammstangi	530	0.0534	88740	4739	4
Hvammstangi	531	0.0427	86080	3672	5
Blönduós	540	0.0455	88740	4035	4
Blönduós	541	0.0442	86080	3804	5
Skagaströnd	545	0.0461	88740	4092	4
Sauðárkrókur	550	0.0540	88740	4794	4
Sauðárkrókur	551	0.0327	86080	2814	5
Varmahlíð	560	0.0424	88740	3760	4
Hofsós	565	0.0487	88740	4325	4
Hofsós	566	0.0642	86080	5523	5
Fljót	570	0.0552	86729	4791	6
Siglufjörður	580	0.0367	86729	3183	4
Akureyri	600	0.0533	86079	4585	3
Akureyri	601	0.0454	86080	3908	5
Akureyri	603	0.0607	86079	5225	3
Grenivík	610	0.0644	88740	5715	4
Grímsey	611	0.1183	88740	10497	4
Dalvík	620	0.0368	88740	3264	4
Dalvík	621	0.0307	86080	2640	5
Ólafsfjörður	625	0.0582	88740	5165	4
Hrísey	630	0.0630	88740	5591	4
Húsavík	640	0.0517	88740	4587	4
Húsavík	641	0.0488	86729	4230	6
Fosshóll	645	0.0559	86080	4814	5
Laugar	650	0.0568	88740	5043	4
Mývatn	660	0.0433	88740	3843	4

Continued on next page

Table 5.2 – continued from previous page

Region	Postal code	Claim frequency	Claim size	Expected compensation	Numeric code
Kópasker	670	0.0465	88740	4131	4
Kópasker	671	0.0389	86729	3371	6
Raufarhöfn	675	0.0428	88740	3797	4
Þórshöfn	680	0.0400	88740	3551	4
Þórshöfn	681	0.0358	86729	3109	6
Bakkafjörður	685	0.0655	88740	5813	4
Vopnafjörður	690	0.0419	88740	3719	4
Egilsstaðir	700	0.0473	86079	4068	3
Egilsstaðir	701	0.0288	86080	2479	5
Seyðisfjörður	710	0.0395	88740	3503	4
Mjóafjörður	715	0.0642	86729	5569	6
Borgarfjörður eystri	720	0.0442	88740	3920	4
Reyðarfjörður	730	0.0429	88740	3803	4
Eskifjörður	735	0.0464	88740	4120	4
Neskaupsstaður	740	0.0418	88740	3712	4
Fáskrúðsfjörður	750	0.0378	88740	3356	4
Stöðvarfjörður	755	0.0398	88740	3530	4
Breiðdalsvík	760	0.0564	88740	5008	4
Djúpivogur	765	0.0397	88740	3519	4
Höfn	780	0.0502	88740	4453	4
Höfn	781	0.0323	86080	2784	5
Öræfi	785	0.0377	86080	3249	5
Selfoss	800	0.0635	86079	5469	3
Selfoss	801	0.0483	86080	4161	5
Hveragerði	810	0.0629	86079	5416	3
Þorlákshöfn	815	0.0515	88740	4572	4
Eyrarbakki	820	0.0638	88740	5664	4
Stokkseyri	825	0.0579	88740	5134	4
Laugarvatn	840	0.0528	86080	4542	5
Flúðir	845	0.0509	88740	4517	4
Hella	850	0.0489	88740	4339	4
Hella	851	0.0395	86080	3397	5
Hvolsvöllur	860	0.0471	88740	4184	4
Hvolsvöllur	861	0.0395	86080	3399	5
Vík	870	0.0647	88740	5737	4
Vík	871	0.0548	86080	4716	5
Kirkjubæjarklaustur	880	0.0516	88740	4583	4
Vestmannaeyjar	900	0.0491	88740	4354	4

Continued on next page

Table 5.2 – continued from previous page

Region	Postal code	Claim frequency	Claim size	Expected compensation	Numeric code
	999	0.0751	88912	6674	8

5.2. Claim frequency

The analysis of claim frequency is based on the Bayesian approach and MCMC. Statistical results are found by using simulated samples to compute the posterior mean, standard deviation and percentiles for all parameters. MCMC trace plots for all parameters are examined and residuals are analyzed. Then model comparison was made with DIC calculations, introduced in Chapter 2. The model was tested with different coefficient C, different postal category matrix and finally models with and without spatial and temporal factors were compared, and the best model chosen. The trace plots for all model parameters are presented in Figures 5.2 - 5.5. There are 16 different a_1 parameters depending on each year. Similarly, there are 130 different postal codes and therefore 130 different a_2 parameters, 8 different regression parameters β according to postal categories and 1974 θ^* parameters depending on the whole data for claim frequency with some number of policy years. For each of a_1, a_2, β and θ^*, one element is picked randomly to show its trace plot. These plots are used to see if all parameters are stable, and Figures 5.2 - 5.5 show stable parameters. It is not necessary to have equal axis for all parameters since there is no need to compare the plots between parameters.

Residual plots based on residual analysis is shown in Figure 5.6. Formula for the residuals is presented in Chapter 4, see Equation (4.2). The figure shows that the residuals for claim frequency behave as expected with the exception that the variance appears to be a little bit higher when expected value is between zero and five constant as opposed to being constant. The average is around zero as expected.

Next step is to compare the for-mentioned model for claim frequency using DIC, the lower DIC value and effective parameters (p_D) give, the better results can be obtained. First, the model was run with different multiplication factor C (introduced in Chapter 4).

Table 5.3 shows, from $C = 4$, the value on effective parameter increases with lower value on C. It also increases when $C = 5$. The model gives good results when $C = 4$ in terms of p_D and DIC is low. Based on these results, a decision is made

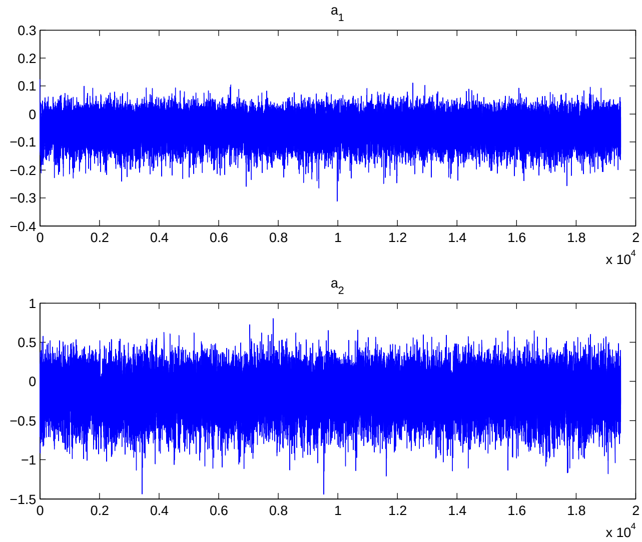

Figure 5.2: MCMC trace plots for parameters $a_{1,5}$ and $a_{2,72}$ for 15000 iterations, burn-in of 4500 and 4 chains.

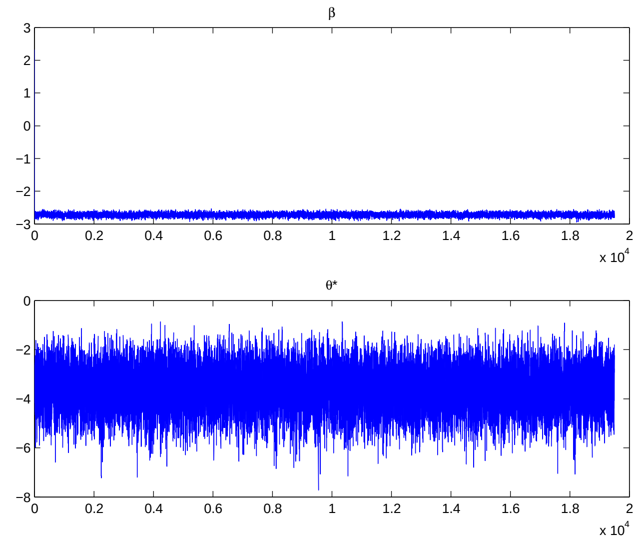

Figure 5.3: MCMC trace plots for parameters β_4 and θ^*_{135} for 15000 iterations, burn-in of 4500 and 4 chains.

to set the value of C equal to 4. The next step is to test the model with different postal category matrices, see results in Table 5.4. First there is only 1 category where all postal codes are in one and the same category. Next there are 6 categories which is similar to the categories introduced in Table 4.1, except postal codes in Reykjavík and Reykjavík urban region are combined and rural areas and rural areas near highway were combined. Then the postal category matrix with eight categories is tried out. Finally the postal categories are 9, which is similar to the 8-th category matrix but there are two categories with small urban region, one close to the highway

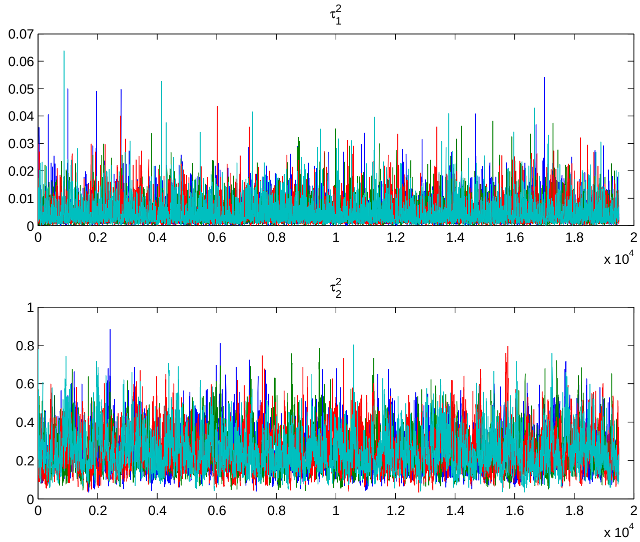

Figure 5.4: MCMC trace plots for parameters τ_1^2 and τ_2^2 for 15000 iterations, burn-in of 4500 and 4 chains.

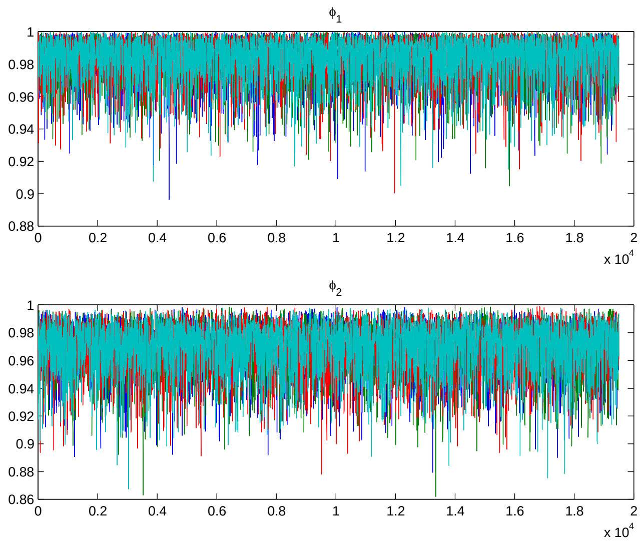

Figure 5.5: MCMC trace plots for parameters ϕ_1 and ϕ_2 for 15000 iterations, burn-in of 4500 and 4 chains.

and the other one not.

Table 5.4 shows the lowest value of both DIC and effective number of parameters is when there are 8 postal categories, which is used in the following calculations. Finally the model is tested with and without the temporal effects and the spatial effects. First, only a_1 is taken out of the model and a_2 is included. Next it is the opposite, a_1 in and a_2 out. Following both parameters are taken out of the model and finally both are included in the model (which should give the same results as

Table 5.3: DIC calculations for Poisson model with different coefficient C. p_D is the effective number of parameters.

DIC calculations		
Coefficient	DIC	p_D
$C = 0.5$	5387.7	589.59
$C = 1$	5381.0	587.48
$C = 1.5$	5378.1	586.77
$C = 2$	5377.5	586.58
$C = 3$	5376.1	586.08
$C = 4$	5375.4	585.78
$C = 5$	5375.6	586.14

Table 5.4: DIC calculations for Poisson model with different postal category matrix. p_D is the effective number of parameters.

DIC calculations		
Postal category matrix	DIC	p_D
X:1	5401.3	599.43
X:6	5377.6	588.42
X:8	5375.4	585.78
X:9	5376.9	586.04

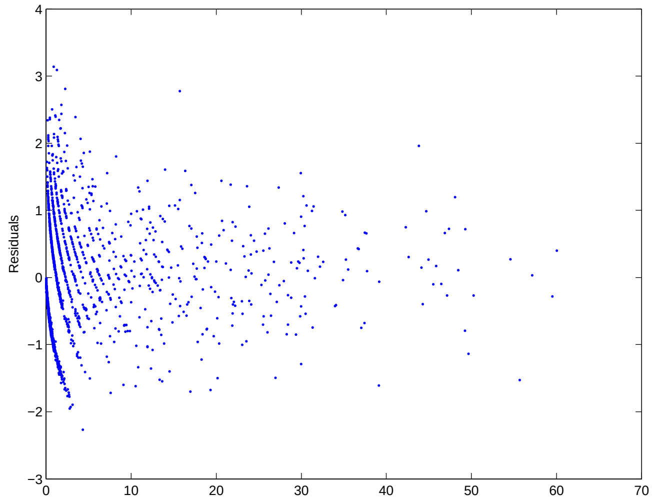

Figure 5.6: Residual plot with residuals of the Poisson model on the y-axis versus $\exp(\mu + 0.5 * \sigma_{it}^2)$ *where* $\mu = X\beta + a_{1,t} + a_{2,i} + \log(e_{it})$ *on the x-axis.*

Table 5.4 with 8 postal categories.

Table 5.5: DIC calculations for Poisson model with and without spatial and temporal effects. p_D *is the effective number of parameters.*

DIC calculations		
Factors	DIC	p_D
a_1 out	5379.0	585.17
a_2 out	5403.1	579.32
a_1 and a_2 out	5410.7	579.23
a_1 and a_2 included	5375.6	585.72

Table 5.5 shows the model is better when a_2 is included. DIC is lower when a_1 and a_2 are both included, but for effective parameters the value is a little bit lower when a_1 is out. According to Figure 5.7 the time factor, $\exp(a_1)$, could be insignificant since the medium line (around 1) does not cut the confidence interval. But Table 5.5 shows the inclusion of a_1 at least dose not make the fit worse. There is a possibility of years with more claims and from Figure 5.7 can be seen that one year, in 1999, number of claims are notably topping, even though the data have been brought to present worth. Also, the lower confidence interval is very close to the medium line so it has been decided for the claim frequency model to include both a_1 and a_2. Then the final model has multiplication factor $C = 4$, 8 postal categories and both a_1 and a_2 included in the model. According to the insurance company, one of a possible reasons why years 1999 and 2000 were heavy in claims is that this was a time of economic growth in the society, which leads to increasing number of cars and more

stressful traffic. After the year 2000 the claim frequency gets lower again which can be because of regression after the growth, discussions about traffic accidents along with consequences of higher premiums, which raised a lot around the year 2000.

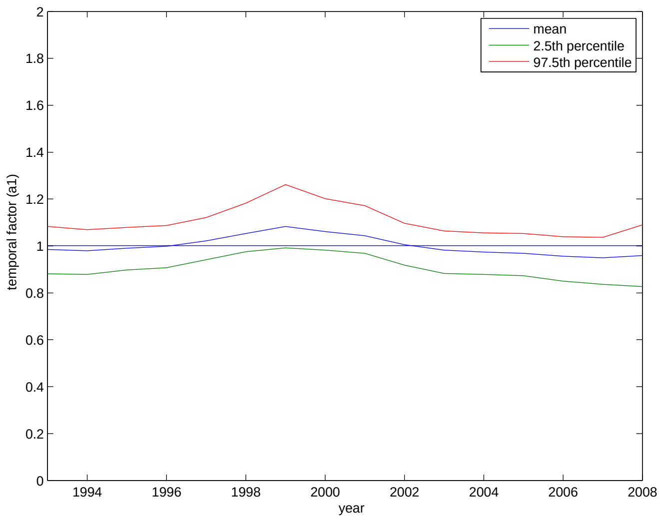

Figure 5.7: The temporal factor a_1 as a function of time. This figure shows the exponential value of the factor, i.e. $\exp(a_1)$.

Following, the results from the Poisson model are analyzed graphically where postal codes are on the x axis and frequency is on the y axis.

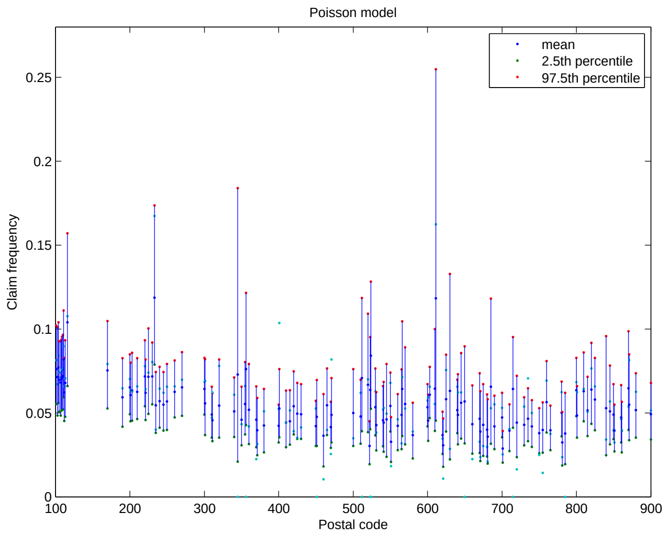

Figure 5.8: Expected claim frequency based on the Poisson model and raw frequency estimates for all postal codes along with 95% posterior interval. The postal codes are on the x-axis and the claim frequency are on the y-axis.

Figure 5.8 shows there is areal trend in the data. It shows the frequency of claims depends on where policyholder is residing. Figures E.1-E.10 in Appendix E give more accurate results where only a part of the postal codes are on each figure. The frequency of claims is highest in the capital area, mostly postal codes that belong to the postal categories Reykjavík and Reykjavík urban region. A few postal codes belonging to the postal category large urban region have similar claim frequency or between 0.06 and 0.08 claims per policy year. Most postal codes in the postal categories small urban regions and rural areas have claim frequency between 0.04 and 0.06 claims per policy year. As can be seen in Figure 5.8, the posterior interval for some postal codes is large. These postal codes are mainly 233, 345 and 611. The reason for this is the same as explained for Figure 5.1, i.e. these postal codes have small amount of claims. During some years, unusual amount of claims occurred when compared to other years for that postal code (the amount of policyholders is still similar) and the frequency of claims in nearby postal codes.

5.3. Claim size

The model for claim size was handled and applied similar by the model for claim frequency. Mean, standard deviation and percentiles were calculated for all parameters. MCMC trace- and residual plots analyzed and model comparison with DIC calculations. Trace plots are shown in Figures 5.9 - 5.11. These plots are used to see if model parameters are stable, and from the figures can be seen all parameters are stable. Like for a_1 and a_2, there are 16 different d_1 parameters and 130 different parameters for d_2. Also 8 different regression parameters η exist according to postal categories. For each of the parameter vectors d_1, d_2 and η, one element is picked randomly to show on the trace plot.

Following, residual plots based on residual analysis are examined. Calculations for the residuals are according to Equation (4.8) introduced in Chapter 4.

Figure 5.12 show that the residuals for the claim size behave like expected. The average is around zero, points equally distributed for different values of the expected claim size which is on the x-axis and no trend in the plot. The variance appears to be constant as a function of expected claim size. Continuing, model comparison was done for changes in the model. Like before, the lower value DIC and effective parameters give, the better results are obtained. Similar to claim frequency, the model is run for different coefficient C, changed postal category matrix and with and without d_1 and d_2.

According to Table 5.6, the value for DIC is very similar independent of coefficient C. But it is obvious the number of effective parameters decreases with lower value on C. Since different values of C only moderately influence the number of effective parameter, p_D in the model, the best result is obtained with $C = 0.5$, which is used

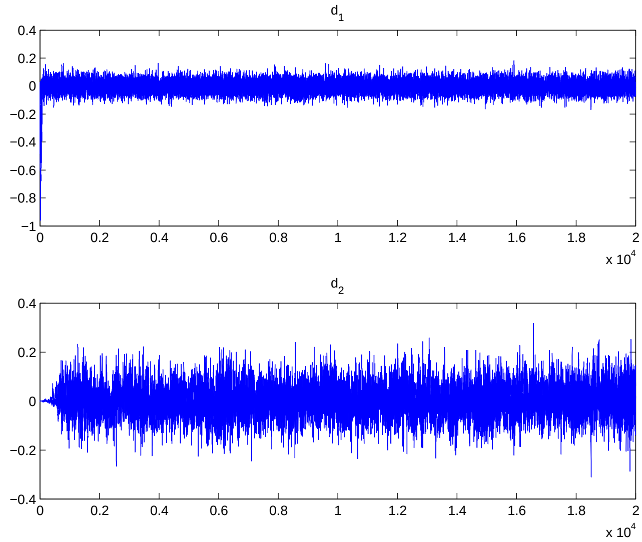

Figure 5.9: MCMC trace plots for parameters $d_{1,11}$ and $d_{2,72}$ for 15000 iterations, burn-in of 5000 and 4 chains.

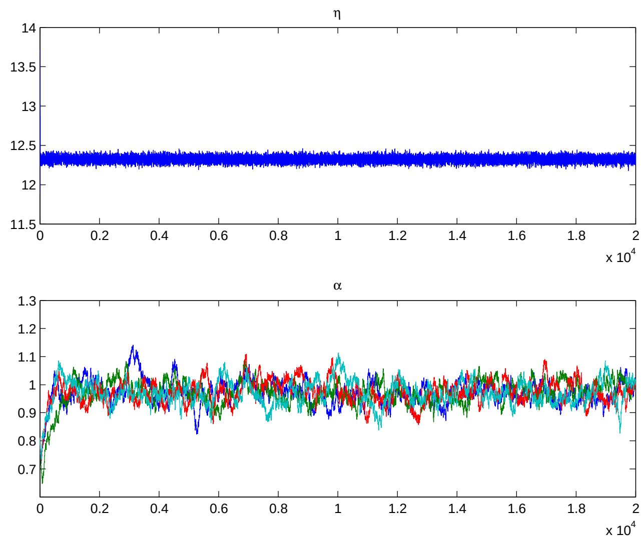

Figure 5.10: MCMC trace plots for parameters α and η_3 for 15000 iterations, burn-in of 5000 and 4 chains.

in the following calculations. Next step is to compare the model with different postal category matrix. The matrix was changed like the model for claim frequency was tested, using 6, 8 and 9 categories, see Subsection 5.2.

The value of DIC and effective number of parameters is very similar independent of size of postal category matrix. The lowest DIC is obtained when there are 9 categories but then the highest number of p_D follows. The lowest p_D is when there are 6 categories but then the highest value of DIC occurs. When there are 8 categories

Table 5.6: DIC calculations for gamma model with different coefficient C. p_D is the effective number of parameters.

DIC calculations		
Coefficient	DIC	p_D
$C = 0.5$	31650	15.55
$C = 1$	31651	16.39
$C = 1.5$	31651	16.85
$C = 2$	31651	17.45
$C = 3$	31651	17.52
$C = 4$	31651	17.73

Table 5.7: DIC calculations for the gamma model with different postal category matrix. p_D is the effective number of parameters.

DIC calculations		
Postal category matrix	DIC	p_D
X:6	31651	14.50
X:8	31650	15.63
X:9	31649	16.13

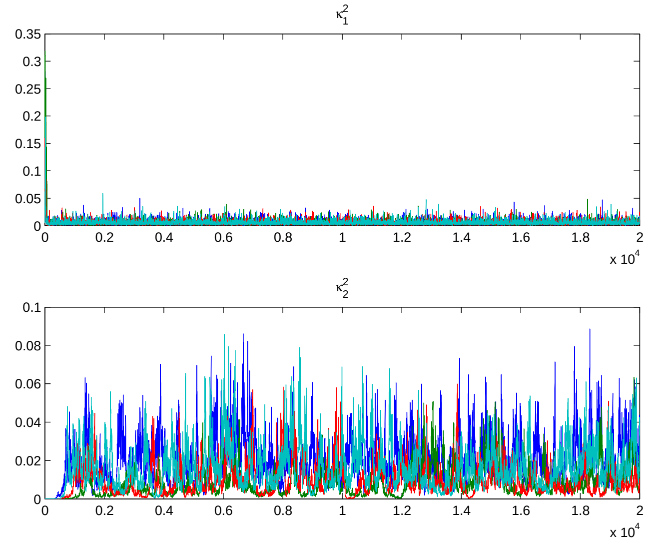

Figure 5.11: MCMC trace plots for parameters κ_1^2 and κ_2^2 for 15000 iterations, burn-in of 5000 and 4 chains.

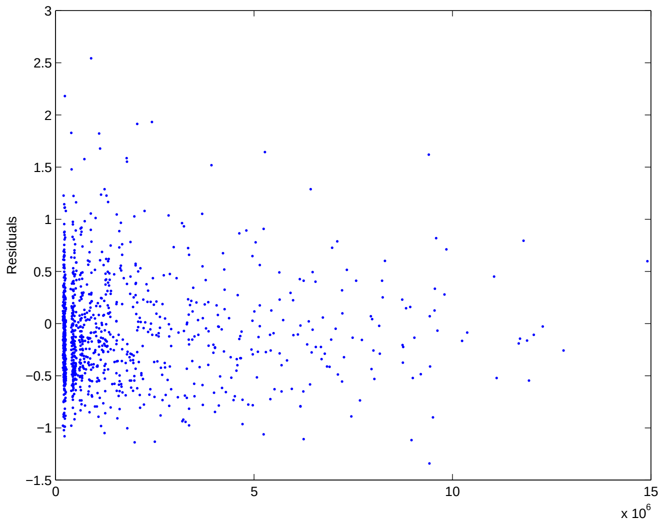

Figure 5.12: Residual plot with residuals of the gamma model on the y-axis versus $N_{it}\exp(\mu^*)$ where $\mu^* = X\eta + d_{1,t}$.

it is in the middle. Since these different categories give very similar results, the model with 8 postal categories is chosen for convenience, because it is the same postal category matrix as used for the claim frequency model and is this result used in following calculations. Finally the model is compared with and without spatial factors.

Table 5.8 shows the model is better when d_1 is included in the model. Results are similar with only d_1 in the model and both d_1 and d_2, whether DIC or the effective

Table 5.8: DIC calculations for gamma model with and without spatial and temporal effects. p_D is the effective number of parameters.

DIC		
Factors	DIC	p_D
d_1 out	31677	34.14
d_2 out	31650	15.63
d_1 and d_2 out	31685	6.41
d_1 and d_2 included	31644	34.16

number of parameters are examined. DIC is though a little bit lower when both factors are included, but after looking more thoroughly at the regional factor d_2, see Figure 5.14, it shows this factor has small influences on the results so it was taken out of the model. But there is always some regional effect because of the postal categories, i.e. d_2 has almost no effect but $X\eta$ has a small effect. According to both Table 5.8 and Figure 5.13 temporal factor has great influences on the results. The figure shows that in 1999 more claims occurs which is in context with information from VÍS that say 1999 was very high in claims. Similar to the claim frequency, the total claim size increases in 1999 and 2000, among others because of economic growth in the society. In times like that, more expensive cars are on the streets and repair cost also raises which leads to higher claims for the insurance company to pay. Also, in 1999 the laws of Tort Damages Act were changed which increased the size of claims, especially for bodily injury, but it also influences the loss of properties. Figure 5.13 shows that the medium line (around 1) is above and below the confidence intervals. It shows even though all data have been brought to present worth, time still has great effects so the present worth does not cover the time effects. So the final model for claim sizes has multiplication factor $C = 0.5$, 8 postal categories and only d_1 included in the model.

Like for the model for claim frequency, spatial trend in the claim sizes will be examined by analyzing the results from the gamma model graphically, see Figure 5.14. This figure shows that the residence of policyholders has a small effect on the total claim sizes. The total claim sizes are between 80000 and 90000. This corresponds to the result in Table 5.8 which shows the spatial factor has almost no effect. But like before the postal category matrix, η has small effect. More accurate pictures will not be shown for this model since Figure 5.14 shows results accurate enough.

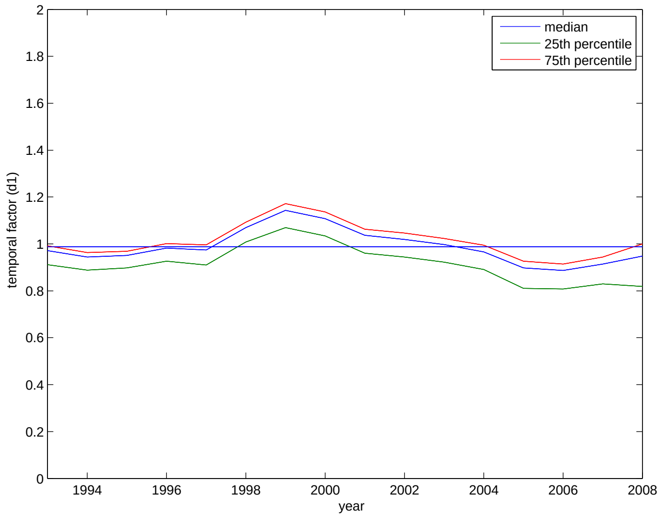

Figure 5.13: The temporal factor d_1 as a function of time. This figure shows the exponential value of the factor, i.e. $\exp(d_1)$.

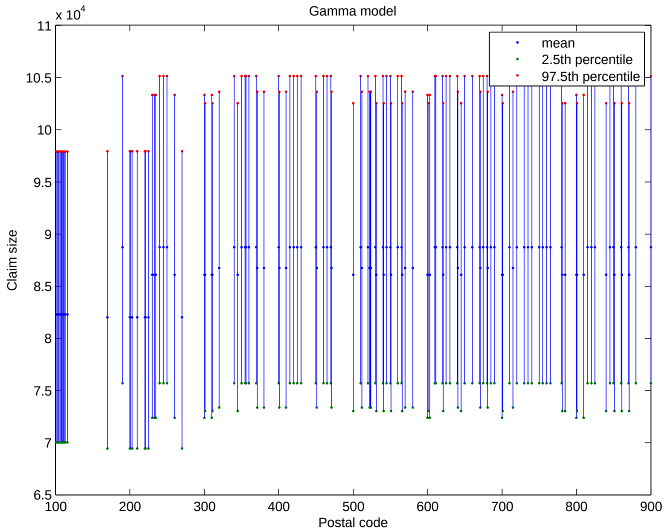

Figure 5.14: Expected total claim size for all postal codes with 95% posterior interval. The postal codes are on the x-axis and the claim size are on the y-axis.

5.4. Expected compensation

Expected compensation cost, introduced in Chapter 4, is based on the combination of results from the Poisson and the gamma models. The expected compensation is the amount the insurance company has to pay per policy year for each postal code in the country.

Like Figure 5.15 shows, the expected compensation is in context with claim fre-

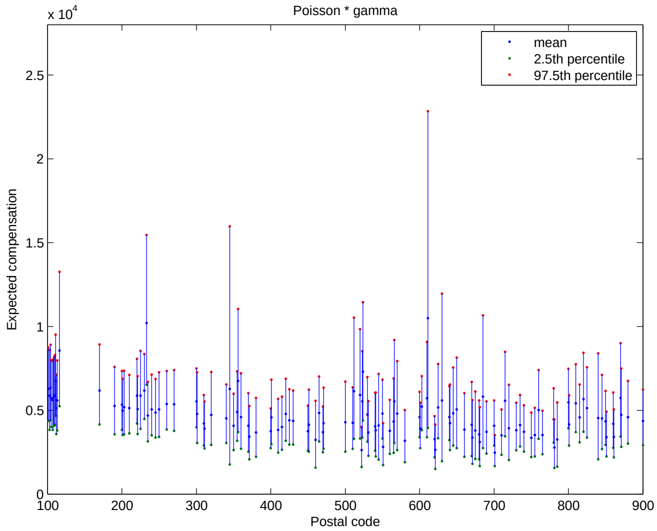

Figure 5.15: Expected compensation cost for all postal codes along with 95% posterior interval.

quency and claim sizes. The total compensation cost depends on residence of policyholder. The highest value is in the capital area and gets lower when you get to the countryside. Figures E.11-E.20 in Appendix E show more accurate results. Figure E.11 and E.12 show the expected cost is around 6000, but in Figure E.13 it starts to get lower and in Figure E.15 the expected cost is stable around 4000. It stays stable except in few cases where the cost raises a bit in postal codes belonging to the postal category Large urban regions and becomes closer to 6000. As mentioned in Section 5.2 there are a few postal codes that have large posterior interval. Those postal codes usually have small amount of claims, but during some years the number of claims rises (but still amount of policyholders similar) and is higher compared to other years and nearby postal codes.

6. Discussion

The aim of this study is to find the expected compensation cost the insurance company needs to pay because of property loss due to car accidents. The information given are based on postal codes and the expected compensation cost is estimated for each postal code in the country. The main goal is to evaluate if the residence of policyholders influence the result, by using a regional factor. Also it was examined whether a temporal factor would improve the model. To improve the regional factor, postal codes were grouped into eight categories depending on type and results examined for each category. In this thesis two kind of neighbor structures are used in the proposed models, one for time and other for postal codes.

Today the insurance company divides the postal codes in the country into a few risk zones. The insurance premium depends, among other factors, on these zones. The matrix describing the neighbor structure is such that it is easy to change if needed. The same holds true for the postal category matrix. The postal category matrix allows both changes on the number of categories and changes of postal codes within each category. Therefore it is easy for the insurance company to change the zones and postal codes within each zone if needed.

The expected compensation cost is based on models for the expected claim frequency and expected total claim size. Results from these two models are combined to get the expected compensation cost. The results from the research can be summarized as follows

- Claim frequency
  - Dependence between claim frequency and the residence of policyholders.
  - Highest number at the capital area and lowest in the country side.
  - Includes both a temporal factor and a regional factor where the regional factor has a great influence but the temporal factor has a small influence.
  - In accordance to DIC comparison the temporal factor is included in case of years with high claim frequency. It does not make the model less qualified.

- Claim size
    - Claim size depends only on postal category but not on individual postal codes.
    - Claim size is similar for the postal categories.
    - Includes a temporal factor.
    - In accordance to DIC comparison the spatial factor (after taking postal categories into account) has little influence and was removed from the model. The temporal factor has great influence despite all data have been brought to present worth.
- Expected compensation
    - In context with the results of claim frequency and claim size.
    - Highest cost in the capital area, lower in the countryside.

Results from the eight postal categories are in context with the results from the models. The expected compensation cost is highest for category 1 which are postal codes in Reykjavík. The cost is a bit lower for category 2 and 3 or Reykjavík urban regions and large urban regions. It gets lower and in small urban regions and rural areas the expected compensation is about 50% lower than in Reykjavík. What makes interest is that the expected compensation cost is a bit higher in rural areas than rural near highway but that is mainly due to higher claim size. For the expected compensation cost the average lower bound multiplication factor is 0.6341 and the average upper bound multiplication factor is 1.5083 which gives an idea about the precision of the expected compensation cost.

According to these results there is a spatial trend in the expected compensation cost so it can be justified to price car insurances according to the residence of the policyholder. The results of this study show the zone separation like the insurance company is using today is not out-of-date but the postal codes can be rearranged in the zones according to the results.

7. Future studies

When working on this thesis, ideas on future projects came up. The main ideas are:

- Further development of the postal categories based on results for each postal code.
- Use a Poisson model without overdispersion.
- Use zero-inflated Poisson model.
- Investigate overdispersed Poisson model where $\exp(\varepsilon_{it})$ follows a gamma distribution.
- For more accurate data it would be interesting to analyze data on individual basis and take into account variables such as
  - Type of car
  - Gender of policyholder
  - Age of policyholder
  - Age of car
  - Accident history of policyholder

  to name a few.
- Analysis data for risk categories with few policyholders.

Bibliography

[1] Gelman A., J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis.* Chapman and Hall/CRC, 2 edition, 2003.

[2] Gelfand A.E. and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *The American Statistician*, 85, 1990.

[3] Spiegelhalter D., N. Best, B. Carlin, and A. van der Linde. Modelling the claims process in the presence of covariates. *J. R. Statist. Soc B*, 64 (4), 2002.

[4] Anscombe F.J. Contribution to the discussion of h. hotelling's paper. *Journal of the Royal Statistical Society, B*, Vol.15,pp.229-230, 1953.

[5] Cramér H. *Mathematical Methods of Statistics.* Prinseton University Press, Princeton, 1946.

[6] Birgir Hrafnkelsson. 08.23.67 selected topics in time series, 2006. Lecture notes.

[7] Birgir Hrafnkelsson. 08.23.69 industrial statistics ii, 2006. Lecture notes.

[8] Migon H.S. and F.A.S. Moura. Hierarchical bayesian collective risk model: an application to health insurance. *Science Direct*, 36, April 2005.

[9] Gill J. *Bayesian Methods A Social and Behavioral Sciences Approach.* Chapman and Hall/CRC, 1 edition, 2002.

[10] The MathWorks. http://www.mathworks.com/.

[11] Metropolis N, A.W. Rosenbluth, and A.H. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1953.

[12] Embrect P., C. Kluppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, 1997.

[13] Kass R.E. and Raftery A.E. Bayes factors. *Journal of the American Statistical Association*, Vol.90,pp.773-795, Jun 1995.

[14] Kass R.E., B.P. Carlin, and R.M. Neal. Markov chain monte carlo in practice: A roundtable discussion. *The American Statistician*, 52, May 1998.

[15] Geman S. and D. Geman. *Stochastic relaxation, Gibbs.* Chapman and Hall/CRC, 1 edition, 2002.

[16] Geschlößl S. and C. Czado. Spatial modelling of claim frequency and claim size in non-life insurance. April 2007.

[17] Gilks W., S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, 1 edition, 1996.

[18] Wikipedia. http://en.wikipedia.org/.

[19] Hastings W.K. Monte carlo sampling methods using markov chains and their applications. *Biometrica*, Vol.57,pp.97-109, 1970.

[20] Dimakos X.K. and A.F. Di Rattalma. Bayesian premium rating with latent structure. *Scand. Actuarial J.*, 2002.

A. Neighborhood structures

A.1. Temporal neighborhood structure

Table A.1: Temporal neighborhood matrix.

Temporal neighborhood matrix																
	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1993	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1994	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1995	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
1996	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
1997	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
1998	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
1999	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
2000	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
2001	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
2002	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
2003	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
2004	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
2005	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
2006	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
2007	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
2008	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

A.2. Regional neighborhood structure

The matrix for regional neighborhood is $N \times N$ matrix, where N is the total number of postal codes. Because of its size only a matrix which shows the neighbors to each postal code will be expressed to describe the neighbors. Instead of 0 and 1 that connects neighbors, the actual postal codes are shown, see Table A.2. Nonetheless the original regional neighborhood matrix is similar to the temporal neighborhood matrix in Table A.1.

Table A.2: Regional neighbors.

Regional neighbours															
Postal code	Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14
101:	Reykjavík	103	104	105	107	108	170								
103:	Reykjavík	101	104	105	108	109	110	200							
104:	Reykjavík	101	103	105	108	109	110	111	112	113	201	203			
105:	Reykjavík	101	103	104	107	108	109	110	200						
107:	Reykjavík	101	105	108	170										
108:	Reykjavík	101	103	104	105	107	109	110	111	200	201	203			
109:	Reykjavík	103	104	105	108	110	111	112	113	200	201	203	210		
110:	Reykjavík	103	104	105	108	109	111	112	113	200	201	203	270		
111:	Reykjavík	104	108	109	110	112	113	200	201	203	210				
112:	Reykjavík	104	109	110	111	113	270								
113:	Reykjavík	104	109	110	111	112	203	270							
116:	Reykjavík	270	300	301	311										
150:	Governments														
155:	Banks														
170:	Seltjarnarnes	101	107												
190:	Vogar	220	221	230	233	235	240	245	260						
200:	Kópavogur	103	105	108	109	110	111	201	203	210					
201:	Kópavogur	104	108	109	110	111	200	203	210	220	221	225			
203:	Kópavogur	104	108	109	110	111	113	200	201	210	220	221	225		
210:	Garðabær	109	111	200	201	203	220	221	225						
220:	Hafnarfjörður	190	201	203	210	221	225	230	233	235	240	260			
221:	Hafnarfjörður	190	201	203	210	220	225	230	233	235	240	260			
225:	Álftanes	201	203	210	220	221									

Continued on next page

Table A.2 – continued from previous page

Regional neighbours															
Postal code	Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14
230:	Reykjanesbær	190	220	221	233	235	240	245	250	260					
233:	Reykjanesbær	190	220	221	230	235	240	245	250	260					
235:	Reykjanesbær	190	220	221	230	233	240	245	250	260					
240:	Grindavík	190	220	221	230	233	235	240	245	250	260				
245:	Sandgerði	190	230	233	235	240	250	260							
250:	Garður	230	233	235	240	245	260								
260:	Reykjanesbær	190	220	221	230	233	235	240	245	250					
270:	Mosfellsbær	110	112	113	116										
300:	Akranes	116	301	310	311	320									
301:	Akranes	116	300	310	311	320									
310:	Borgarnes	300	301	311	320	370	371	530	531						
311:	Borgarnes	116	300	301	310	320	340	350	355	356	360	370	371	530	531
320:	Reykholt	300	301	310	311	370	371	530	531						
340:	Stykkishólmur	311	345	350	355	356	360	370	371						
345:	Flatey á Breiðarfirði	340													
350:	Grundarfjörður	311	340	355	356	360	370	371							
355:	Ólafsvík	311	340	350	356	360									
356:	Snæfellsbær	311	340	350	355	360									
360:	Hellissandur	311	340	350	355	356									
370:	Búðardalur	310	311	320	340	350	371	380	500						
371:	Búðardalur	310	311	320	340	350	370	380	500						
380:	Reykhólahreppur	370	371	420	450	451	465	500							
400:	Ísafjörður	401	410	415	420	425	430	450	451	465	470	471	522	523	524
401:	Ísafjörður	400	410	415	420	425	430	450	451	465	470	471	522	523	524

Continued on next page

Table A.2 – continued from previous page

		Regional neighbours													
Postal code	Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14
410:	Hnífsdalur	400	401	415	420	425	430	450	451	465	470	471	522	523	524
415:	Bolungarvík	400	401	410	420	425	430	470	471						
420:	Súðavík	380	400	401	410	415	425	430	450	451	465	470	471		
425:	Flateyri	400	401	410	415	420	430	450	451	465	470	471	522	523	524
430:	Suðureyri	400	401	410	415	420	425	450	451	465	470	471	522	523	524
450:	Patreksfjörður	380	400	401	410	420	425	430	451	460	465	470	471		
451:	Patreksfjörður	380	400	401	410	420	425	430	450	460	465	470	471		
460:	Tálknafjörður	450	451	465											
465:	Bíldudalur	380	400	401	410	420	425	430	450	451	460	470	471		
470:	Þingeyri	400	401	410	415	420	425	430	450	451	465	471	522	523	524
471:	Þingeyri	400	401	410	415	420	425	430	450	451	465	470	522	523	524
500:	Staður	370	371	380	510	530	531								
510:	Hólmavík	500	512	520	522	523	524								
512:	Hólmavík	510	520	522	523	524									
520:	Drangsnes	510	512	522	523	524									
522:		400	401	410	425	430	470	471	510	512	520	523	524		
523:		400	401	410	425	430	470	471	510	512	520	522	524		
524:	Árneshreppur	400	401	410	425	430	470	471	510	512	520	522	523		
530:	Hvammstangi	310	311	320	500	531	540	541							
531:	Hvammstangi	310	311	320	500	530	540	541							
540:	Blönduós	530	531	541	545	551									
541:	Blönduós	530	531	540	551										
545:	Skagaströnd	540													
550:	Sauðárkrókur	551	560	565	566	570									

Continued on next page

Table A.2 – continued from previous page

		Regional neighbours													
Postal code	Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14
551:	Sauðárkrókur	540	541	550	560	565	566	570	580						
560:	Varmahlíð	550	551	565	566	570									
565:	Hofsós	550	551	560	566	570									
566:	Hofsós	550	551	560	565	570									
570:	Fljót	550	551	560	565	566	580								
580:	Siglufjörður	551	570	625											
600:	Akureyri	601	603	610	620	621									
601:	Akureyri	600	603	610	611	620	621	625	630	641	645	650	660		
603:	Akureyri	600	601	610	620	621									
610:	Grenivík	600	601	603	641										
611:	Grímsey	601													
620:	Dalvík	600	601	603	621	625									
621:	Dalvík	600	601	603	620	625									
625:	Ólafsfjörður	580	601	620	621										
630:	Hrísey	601													
640:	Húsavík	640	641	645	650	660	670	671	675						
641:	Húsavík	601	610	640	645	650	660								
645:	Fosshóll	601	640	641	650	660									
650:	Laugar	601	640	641	645	660									
660:	Mývatn	601	640	641	645	650									
670:	Kópasker	640	671	675	680	681									
671:	Kópasker	640	670	675	680	681									
675:	Raufarhöfn	640	670	671	680	681									
680:	Þórshöfn	670	671	675	681	685	690								

Continued on next page

Table A.2 – continued from previous page

		Regional neighbours													
Postal code	Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14
681:	Þórshöfn	670	671	675	680	685	690								
685:	Bakkafjörður	680	681	690											
690:	Vopnafjörður	680	681	685	700	701									
700:	Egilsstaðir	690	701	710	715	720	730	735	740	750	755	760	765		
701:	Egilsstaðir	690	700	710	715	720	730	735	740	750	755	760	765		
710:	Seyðisfjörður	700	701	715	720	730	735	740	750	755					
715:	Mjóafjörður	700	701	710	720	730	735	740	750	755	760	765			
720:	Borgarfjörður eystri	700	701	710	715										
730:	Reyðarfjörður	700	701	710	715	735	740	750	755	760	765				
735:	Eskifjörður	700	701	710	715	730	740	750	755	760	765				
740:	Neskaupsstaður	700	701	710	715	730	735	750	755	760	765				
750:	Fáskrúðsfjörður	700	701	710	715	730	735	740	755	760	765				
755:	Stöðvarfjörður	700	701	710	715	730	735	740	750	760	765				
760:	Breiðdalsvík	700	701	715	730	735	740	750	755	765	780				
765:	Djúpivogur	700	701	715	730	735	740	750	755	760	780	781	785		
780:	Höfn	760	765	781	785	880									
781:	Höfn	765	780	785	880										
785:	Öræfi	765	780	781	880										
800:	Selfoss	801	810	815	820	825	850	851							
801:	Selfoss	800	810	815	820	825	840	845	850	851	860	861			
810:	Hveragerði	800	801	815	820	825									
815:	Þorlákshöfn	800	801	810	820	825	900								
820:	Eyrarbakki	800	801	810	815	825	850	851							
825:	Stokkseyri	800	801	810	815	820	850	851							

Continued on next page

Table A.2 – continued from previous page

		Regional neighbours													
Postal code	Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14
840:	Laugarvatn	801	845												
845:	Flúðir	801	840												
850:	Hella	800	801	801	820	825	851	860	861	870	871	880			
851:	Hella	800	801	801	820	825	850	860	861	870	871	880			
860:	Hvolsvöllur	801	850	851	861	870	871	880							
861:	Hvolsvöllur	801	850	851	860	870	871	880							
870:	Vík	850	851	860	861	871	880								
871:	Vík	850	851	860	861	870	880								
880:	Kirkjubæjarklaustur	780	781	785	850	851	860	861	870	871					
900:	Vestmannaeyjar	815													
999:	Other														

B. Calculations for σ_{it}^2

Approximation for σ_{it}^2

$$N_{it} \sim \text{Poisson}(e_{it} \cdot \text{r}),$$

where r is the average frequency per policy year, $r = \exp(C_1)$ and $C_1 = \log(\theta)$.

$$\hat{r} = \frac{\sum_i \sum_t N_{it}}{\sum_i \sum_t e_{it}}$$

where $\hat{C}_1 = \log(\hat{r})$ and $\hat{C}_2 = \exp(-\hat{C}_1)$.

$$\mu = X^T\beta + a_1 + a_2 + \log e_{it}$$

$X^T\beta \simeq C_1$, $a_1 \simeq 0$ and $a_2 \simeq 0$, which gives

$$\exp(\mu) \simeq e_{it} \exp(C_1)$$

$$\exp(\mu)\exp(2\sigma_{it}^2) - \exp(\mu)\exp(\sigma_{it}^2) + \exp(\sigma_{it}^2/2) = (1+\delta)$$
$$\exp(2\sigma_{it}^2) - \exp(\sigma_{it}^2) + \exp(\sigma_{it}^2/2 - \mu) = (1+\delta)\exp)(\mu)$$
$$\exp(2\sigma_{it}^2) - \exp(\sigma_{it}^2) = (1 + \delta - \exp(\sigma_{it}^2/2))\exp(-\mu)$$

or

$$\exp(2\sigma_{it}^2) - \exp(\sigma_{it}^2) \simeq \exp(-\mu) = \frac{1}{e_{it}}\exp(-C_1)$$

which gives the second order equation

$$\exp(2\sigma_{it}^2) - \exp(\sigma_{it}^2) - \frac{C_2}{e_{it}} = 0.$$

Then

$$\exp(\sigma_{it}^2) = \frac{1 + \sqrt{1 + 4C_2/e_{it}}}{2}$$

and

$$\sigma_{it}^2 = \log\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + 4C_2/e_{it}}\right) = \log\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + 4/(r + e_{it})}\right)$$

C. Conditional distributions

Conditional distributions for parameters in model for claim frequency:

θ^*_{it}, i = i,...,J and t = 1,...,T

$$\begin{aligned} p(\theta^*_{it}|rest) &\propto p(N_{it}|\theta^*_{it}) \times p(\theta^*_{it}|a_1, a_2, \beta, \sigma^2_{\varepsilon,it}) \\ &\propto \text{Poisson}[N_{it}|e_{it}\exp(\theta^*_{it})] \times \text{N}(\theta^*_{it}|\mu_{it}, \sigma^2_{\varepsilon,it}) \\ &\propto e_{it}^{N_{it}} \exp(N_{it}\theta^*_{it}) \exp[-e_{it}\exp(\theta^*_{it})] \exp\left[-\tfrac{1}{2\sigma^2_{\varepsilon,it}}(\theta^*_{it} - \mu_{it})^2\right] \end{aligned}$$

where

$$\mu_{it} = x_i^T\beta + Z_1a_1 + Z_2a_2$$

so

$$\log p(\theta^*_{it}|rest) = C_0 + N_{it}\log(e_{it}) + N_{it}\theta^*_{it} + \log(e_{it}) - \exp(\theta^*_{it}) - \frac{1}{2\sigma^2_{\varepsilon,it}}(\theta^*_{it} - \mu_{it})^2$$

and the proposal variance is

$$\text{Var}(\theta^*_{it})_{prop} \propto \frac{1}{N_{it} + 1/\sigma^2_{\varepsilon,it}}$$

Conditional distribution of β

$$\begin{aligned} p(\beta|\text{rest}) &\propto p(\theta^*|a_1, a_2, \beta, \Sigma_\varepsilon) \times p(\beta) \\ &\propto \text{N}(\theta^*|\gamma, \Sigma_\varepsilon I) \times \text{N}(\beta|\mu_\beta, \sigma^2_\beta I) \\ &\propto \exp\left[-\tfrac{1}{2}(\theta^* - \gamma)^T\Sigma_\varepsilon^{-1}(\theta^* - \gamma)\right] \exp\left[-\tfrac{1}{2\sigma^2_\beta}(\beta - \mu_\beta)^2\right] \\ \log p(\beta|rest) &= C_0 - \tfrac{1}{2}\left[\beta^TX^T\Sigma_\varepsilon^{-1}X\beta - 2\beta^TX^T(\theta^* - Z_1a_1 - Z_2a_2)\right] - \tfrac{1}{2}(\beta - \mu_\beta)^T\Sigma_\beta^{-1}(\beta - \mu_\beta) \end{aligned}$$

where

$$\gamma = X\beta + Z_1a_1 + Z_2a_2$$

since

$$p(\beta|rest) = N(\beta|\mu_\beta, \sigma_\beta)$$

the covariance matrix is

$$\Sigma_\beta = (X^T\Sigma_\varepsilon^{-1}X + \sigma_\beta^2 I)^{-1}$$

and the mean is

$$\mu_\beta = \Sigma_\beta(X^T\Sigma_\varepsilon^{-1}(\theta^* - Z1a1 - Z2a2) + \sigma_\beta^{-2}\mu_\beta)$$

The conditional distribution of a_1

$$\begin{aligned}
p(a_1|\text{rest}) &\propto p(\theta^*|a_1, a_2, \beta, \Sigma_\varepsilon) \times p(a_1|\phi_1, \tau_1^2) \\
&\propto \mathrm{N}(\theta^*|\gamma, \Sigma_\varepsilon I) \times \mathrm{N}(a_1|\bar{0}, \tau_1^2(I - \phi_1 C_1)^{-1}M_1) \\
&\propto \exp\left[-\tfrac{1}{2}(\theta^* - \gamma)^T\Sigma_\varepsilon^{-1}(\theta^* - \gamma)\right] \\
&\quad \times \exp\left[-\tfrac{1}{2\tau_1^2}a_1^T(M_1^{-1} - \phi_1 M_1^{-1}C_1)a_1\right] \\
\log p(a_1|\text{rest}) &= C_0 - \tfrac{1}{2}[a_1^T Z_1^T\Sigma_\varepsilon^{-1}Z_1 a_1 - 2a_1^T Z_1^T\Sigma_\varepsilon^{-1}(\theta^* - X\beta - Z_2 a_2)] \\
&\quad -\tfrac{1}{2\tau_1^2}a_1^T(M_1^{-1} - \phi_1 M_1^{-1}C_1)a_1
\end{aligned}$$

where

$$\gamma = X\beta + Z_1 a_1 + Z_2 a_2$$

And since

$$p(a_1|rest) = N(a_1|\mu_{a1}, \Sigma_{a1})$$

the covariance matrix is

$$\Sigma_{a1} = [Z_1^T\Sigma_\varepsilon^{-1}Z_1 + \tau_1^2(M_1^{-1} - \phi_1 M_1^{-1}C_1)]^{-1}$$

and the mean is

$$\mu_{a1} = \Sigma_{a1}Z_1^T\Sigma_\varepsilon^{-1}(\theta^* - X\beta - Z_2 a_2).$$

Then a_1 is adjusted so that $\Sigma_j a_{1,j} = 0$ and $Q^{-1} = \Sigma_{a1}$ so

$$a_1^* = a_1 - Q^{-1}A^T(AQ^{-1}A^T)^{-1}(Aa_1 - 0),$$

where A is a vector of ones.

And similar, the conditional distribution of a_2 is

$$\begin{aligned}
p(a_2|\text{rest}) &\propto p(\theta^*|a_1, a_2, \beta, \Sigma_\varepsilon) \times p(a_2|\phi_2, \tau_2^2) \\
&\propto \mathrm{N}(\theta^*|\gamma, \Sigma_\varepsilon I) \times \mathrm{N}(a_2|\bar{0}, \tau_2^2(I - \phi_2 C_2)^{-1} M_2) \\
&\propto \exp\left[-\tfrac{1}{2}(\theta^* - \gamma)^T \Sigma_\varepsilon^{-1}(\theta^* - \gamma)\right] \\
&\quad \times \exp\left[-\tfrac{1}{2\tau_2^2} a_2^T (M_2^{-1} - \phi_2 M_2^{-1} C_2) a_2\right] \\
\log p(a_2|\text{rest}) &= C_0 - \tfrac{1}{2}[a_2^T Z_2^T \Sigma_\varepsilon^{-1} Z_2 a_2 - 2a_2^T Z_2^T \Sigma_\varepsilon^{-1}(\theta^* - X\beta - Z_1 a_1)] \\
&\quad - \tfrac{1}{2\tau_1^2} a_2^T (M_2^{-1} - \phi_2 M_2^{-1} C_2) a_2
\end{aligned}$$

where

$$\gamma = X\beta + Z_1 a_1 + Z_2 a_2$$

And since

$$p(a_2|rest) = N(a_2|\mu_{a2}, \Sigma_{a2})$$

the covariance matrix is

$$\Sigma_{a2} = [Z_2^T \Sigma_\varepsilon^{-1} Z_2 + \tau_2^2 (M_2^{-1} - \phi_2 M_2^{-1} C_2)]^{-1}$$

and the mean is

$$\mu_{a2} = \Sigma_{a2} Z_2^T \Sigma_\varepsilon^{-1} (\theta^* - X\beta - Z_1 a_1)$$

Then a_2 is adjusted so that $\Sigma_j a_{2,j} = 0$ and $Q^{-1} = \Sigma_{a2}$ so

$$a_2^* = a_2 - Q^{-1} A^T (A Q^{-1} A^T)^{-1} (A a_2 - 0),$$

where A is a vector of ones.

The conditional distribution of τ_1^2:

$$\begin{aligned}
p(\tau_1^2|\text{rest}) &\propto p(a_1|\phi_1, \tau_1^2) \times p(\tau_1^2) \\
&\propto \mathrm{N}[a_1|\bar{0}, \tau_1^2 (I - \phi_1 C_1)^{-1} M_1] \times \text{Inv-}\chi^2(\nu_{\tau 1}, s_{\tau 1}^2) \\
&\propto (\tau_1^2)^{-n_1/2} \exp\left[-\tfrac{1}{2\tau_1^2} a_1^T (M_1^{-1} - \phi_1 M_1^{-1} C_1) a_1\right] (\tau_1^2)^{-(\nu_{\tau 1}/2+1)} \exp\left(-\tfrac{\nu_{\tau 1} s_{\tau 1}^2}{2\tau_1^2}\right) \\
&\propto (\tau_1^2)^{-[(n_1+\nu_{\tau 1})/2+1]} \exp\left\{-\tfrac{1}{2\tau_1^2}[\nu_{\tau 1} s_{\tau 1}^2 + a_1^T (M_1^{-1} - \phi_1 M_1^{-1} C_1) a_1]\right\} \\
&\propto \text{Inv-}\chi^2\{n_1 + \nu_{\tau 1}, (n_1 + \nu_{\tau 1})^{-1}[\nu_{\tau 1} s_{\tau 1}^2 + a_1^T (M_1^{-1} - \phi_1 M_1^{-1} C_1) a_1]\}
\end{aligned}$$

The conditional distribution of ϕ_1:

$$\begin{aligned}
p(\phi_1|\text{rest}) &\propto p(a_1|\phi_1,\tau_1^2) \times p(\phi_1) \\
&\propto \text{N}[a_1|\bar{0}, \tau_1^2(I-\phi_1 C_1)^{-1}M_1] \times \text{beta}(\phi_1|\alpha_{\phi 1}, \beta_{\phi 1}) \\
&\propto |I-\phi_1 C_1|^{1/2} \exp\left[-\tfrac{1}{2\tau_1^2} a_1^T (M_1^{-1} - \phi_1 M_1^{-1} C_1) a_1\right] \times \phi_1^{\alpha_{\phi 1}-1}(1-\phi_1)^{\beta_{\phi 1}-1}
\end{aligned}$$

Let $\lambda_{1(j)}$ be the ordered eigenvalue of C_1, then

$$|I-\phi_1 C_1| = \prod_{j=1}^{n_1}(1-\phi_1\lambda_{1(j)})$$

$$\log p(\phi_1|rest) = C_0 + 0.5\sum_{j=1}^{n_1}\log(1-\phi_1\lambda_{1(j)}) + 0.5\phi_1\tau_1^{-2}a_1^T(M_1^{-1}C_1)a_1$$
$$+(\alpha_{\phi 1}-1)\log(\phi_1) + (\beta_{\phi 1}-1)\log(1-\phi_1)$$

The conditional distribution of τ_2^2:

$$\begin{aligned}
p(\tau_2^2|\text{rest}) &\propto p(a_2|\phi_2,\tau_2^2) \times p(\tau_2^2) \\
&\propto \text{N}(a_2|\bar{0}, \tau_2^2(I-\phi_2 C_2)^{-1}M_2) \times \text{Inv-}\chi^2(\nu_{\tau 2}, s_{\tau 2}^2) \\
&\propto (\tau_2^2)^{-n_2/2} \exp\left[-\tfrac{1}{2\tau_2^2} a_2^T (M_2^{-1} - \phi_2 M_2^{-1} C_2) a_2\right] \times (\tau_2^2)^{-(\nu_{\tau 2}/2+1)} \exp\left(-\tfrac{\nu_{\tau 2} s_{\tau 2}^2}{2\tau_2^2}\right) \\
&\propto (\tau_2^2)^{-[(n_2+\nu_{\tau 2})/2+1]} \exp\left\{-\tfrac{1}{2\tau_2^2}[\nu_{\tau 2} s_v^2 + a_2^T(M_2^{-1} - \phi_2 M_2^{-1} C_2) a_2]\right\} \\
&\propto \text{Inv-}\chi^2\{n_2+\nu_{\tau 2}, (n_2+\nu_{\tau 2})^{-1}[\nu_{\tau 2} s_{\tau 2}^2 + a_2^T(M_2^{-1} - \phi_2 M_2^{-1} C_2) a_2]\}
\end{aligned}$$

The conditional distribution of ϕ_2:

$$\begin{aligned}
p(\phi_2|\text{rest}) &\propto p(a_2|\phi_2,\tau_2^2) \times p(\phi_2) \\
&\propto \text{N}(a_2|\bar{0}, \tau_2^2(I-\phi_2 C_2)^{-1}M_2) \times \text{beta}(\phi_2|\alpha_{\phi 2}, \beta_{\phi 2}) \\
&\propto |I-\phi_2 C_2|^{1/2} \exp\left[-\tfrac{1}{2\tau_2^2} a_2^T (M_2^{-1} - \phi_2 M_2^{-1} C_2) a_2\right] \times \phi_2^{\alpha_{\phi 2}-1}(1-\phi_2)^{\beta_{\phi 2}-1}
\end{aligned}$$

Let $\lambda_{2(j)}$ be the ordered eigenvalue of C_2, then

$$|I-\phi_2 C_2| = \prod_{j=1}^{n_2}(1-\phi_2\lambda_{2(j)})$$

$$\log p(\phi_2|rest) = C_0 + 0.5\sum_{j=1}^{n_2}\log(1-\phi_2\lambda_{2(j)}) + 0.5\phi_2\tau_2^{-2}a_2^T(M_2^{-1}C_2)a_2$$
$$+(\alpha_{\phi 2}-1)\log(\phi_2) + (\beta_{\phi 2}-1)\log(1-\phi_2)$$

Following, conditional distributions for all parameters in the claim size model:

The conditional distribution of η

$$\begin{aligned}
P(\eta|\text{rest}) &\propto p(S|d_1,d_2,\eta,\alpha,N)\times p(\eta) \\
&\propto \textstyle\prod_{i=1}^{J}\prod_{t=1}^{T}\text{gamma}\left[S_{it}|N_{it}\exp(\lambda_{it}),\frac{N_{it}}{\alpha}\exp(2\lambda_{it})\right]\times \text{N}(\eta|\mu_\eta,\Sigma_\eta I) \\
&\propto \textstyle\prod_{i=1}^{J}\prod_{t=1}^{T}[\alpha\exp(-\lambda_{it})]^{\alpha N_{it}}\exp[-\alpha\exp(-\lambda_{it})S_{it}]\times\exp\left[\frac{-1}{2}(\eta-\mu_\eta)^T\Sigma_\eta^{-1}(\eta-\mu_\eta)\right] \\
\log(\eta|\text{rest}) &= \textstyle\sum_{i=1}^{J}\sum_{t=1}^{T}\{(\alpha N_{it})\log[\alpha\exp(-\lambda_{it})]-\alpha\exp(-\lambda_{it})S_{it}\}-\frac{1}{2}(\eta-\mu_\eta)^T\Sigma_\eta^{-1}(\eta-\mu_\eta)
\end{aligned}$$

where
$$\lambda_{it} = x_i^T\eta + d_{1,t} + d_{2,i}$$

And the proposal variance is
$$\text{Var}(\eta)_{prop} \propto \frac{1}{\alpha\sum_{i=1}^{J}\sum_{t=1}^{T}N_{it}x_i} + \frac{1}{\sigma_\eta^2}$$

Conditional distribution of $d_{1,t}$, $t=1,...,T$

$$\begin{aligned}
P(d_{1,t}|\text{rest}) &\propto p(S|d_1,d_2,\eta,\alpha,N)\times p(d_1|\kappa_1^2,\zeta_1) \\
&\propto \textstyle\prod i=1J\text{gamma}\left[S_{it}|N_{it}\exp(\lambda_{it}),\frac{N_{it}}{\alpha}\exp(2\lambda_{it})\right] \\
&\quad \times\text{N}\left(d_{1,t}|\textstyle\sum_{k=1,k\neq t}^{T}C_{1,tk}d_{1,k},M_{1,tt}\kappa_1^2\right) \\
&\propto \textstyle\prod i=1J\alpha\exp(-\lambda_{it})^{\alpha N_{it}}\exp[-\alpha\exp(-\lambda_{it})S_{it}] \\
&\quad \times\exp\left[\frac{-1}{2M_{1,tt}\kappa_1^2}\left(d_{1,t}-\textstyle\sum_{k=1,k\neq t}^{T}C_{1,tk}d_{1,k}\right)^2\right] \\
&\propto \textstyle\prod i=1J\exp(-\alpha N_{it}\lambda_{it})\exp[-\alpha S_{it}\exp(-\lambda_{it})] \\
&\quad \times\exp\left[\frac{-1}{2M_{1,tt}\kappa_1^2}\left(d_{1,t}-\textstyle\sum_{k=1,k\neq t}^{T}C_{1,tk}d_{1,k}\right)^2\right] \\
\log p(d_{1,t}|\text{rest}) &= \textstyle\sum_{i=1}^{J}[-\alpha N_{it}\exp(\lambda_{it})-\alpha S_{it}\exp(-\lambda_{it})]-\frac{-1}{2M_{1,tt}\kappa_1^2}\left(d_{1,t}-\textstyle\sum_{k=1,k\neq t}^{T}C_{1,tk}d_{1,k}\right)^2
\end{aligned}$$

where

$$\lambda_{it} = x_i^T \eta + d_{1,t} + d_{2,i}$$

And the proposal variance is

$$\text{Var}(d_{1,t})_{prop} \propto \frac{1}{\alpha \sum_{i=1}^{J} N_{it} + \frac{1}{M_{1,tt}\kappa_1^2}}$$

Conditional distribution of $d_{2,i}$, $i = 1, ..., J$

$$\begin{aligned}
P(d_{2,i}|\text{rest}) &\propto p(S|d_1, d_2, \eta, \alpha, N) \times p(d_2|\kappa_2^2, \zeta_2) \\
&\propto \prod t = 1T\text{gamma}\left[S_{it}|N_{it}\exp(\lambda_{it}), \tfrac{N_{it}}{\alpha}\exp(2\lambda_{it})\right] \\
&\quad \times \text{N}\left(d_{2,i}| \textstyle\sum_{k=1,k\neq i}^{T} C_{2,ik}d_{2,k}, M_{2,ii}\kappa_2^2\right) \\
&\propto \prod t = 1T\alpha\exp(-\lambda_{it})^{\alpha N_{it}}\exp[-\alpha\exp(-\lambda_{it})S_{it}] \\
&\quad \times \exp\left[\tfrac{-1}{2M_{2,ii}\kappa_2^2}\left(d_{2,i} - \textstyle\sum_{k=1,k\neq i}^{T} C_{2,ik}d_{2,k}\right)^2\right] \\
&\propto \prod t = 1T\exp(-\alpha N_{it}\lambda_{it})\exp[-\alpha S_{it}\exp(-\lambda_{it})] \\
&\quad \times \exp\left[\tfrac{-1}{2M_{2,ii}\kappa_2^2}\left(d_{2,i} - \textstyle\sum_{k=1,k\neq i}^{T} C_{2,ik}d_{2,i}\right)^2\right] \\
\log P(d_{2,i}|\text{rest}) &= \textstyle\sum_{t=1}^{T}[-\alpha N_{it}\exp(\lambda_{it}) - \alpha S_{it}\exp(-\lambda_{it})] - \tfrac{-1}{2M_{2,ii}\kappa_2^2}\left(d_{2,i} - \textstyle\sum_{k=1,k\neq i}^{T} C_{2,ik}d_{2,k}\right)^2
\end{aligned}$$

where

$$\lambda_{it} = x_i^T \eta + d_{1,t} + d_{2,i}$$

And the proposal variance is

$$\text{Var}(d_{2,i})_{prop} \propto \frac{1}{\alpha \sum_{t=1}^{T} N_{it} + \frac{1}{M_{2,ii}\kappa_1^2}}$$

Conditional distribution of κ_1^2

$$\begin{aligned}
P(\kappa_1^2|\text{rest}) &\propto p(d_1|\zeta_1,\kappa_1^2)\times p(\kappa_1^2)\\
&\propto \text{N}[d_1|\bar{0},\kappa_1^2(I-\zeta_1C_1)^{-1}M_1]\times \text{Inv-}\chi^2(\nu_{\kappa 1},S_{\kappa 1}^2)\\
&\propto (\kappa_1^2)^{-n_1/2}\exp\left[\tfrac{-1}{2\kappa_1^2}d_1^T(M_1^{-1}-\zeta_1M_1^{-1}C_1)d_1\right]\times(\kappa_1^2)^{-(\nu_{\kappa_1}/2+1)}\exp\left(-\tfrac{\nu_{\kappa 1}S_{\kappa 1}^2}{2\kappa_1^2}\right)\\
&\propto (\kappa_1^2)^{-\left(\frac{n_1\nu_{\kappa 1}}{2}+1\right)}\exp\left\{\tfrac{-1}{2\kappa_1^2}[\nu_{\kappa 1}S_{\kappa 1}^2+d_1^T(M_1^{-1}-\zeta_1M_1^{-1}C_1)d_1]\right\}\\
&\propto \text{Inv-}\chi^2\{n_1+\nu_{\kappa 1},(n_1+\nu_{\kappa 1})^{-1}[\nu_{\kappa 1}S_{\kappa 1}^2+d_1^T(M_1^{-1}-\zeta_1M_1^{-1}C_1)d_1]\}
\end{aligned}$$

Conditional distribution of ζ_1

$$\begin{aligned}
P(\zeta_1|\text{rest}) &\propto p(d_1|\zeta_1,\kappa_1^2)\times p(\zeta_1)\\
&\propto \text{N}[d_1|\bar{0},\kappa_1^2(I-\zeta_1C_1)^{-1}M_1]\times \text{beta}(\zeta_1|\alpha_{\zeta 1},\beta_{\zeta 1})\\
&\propto |I-\zeta_1C_1|^{1/2}\exp\left[\tfrac{-1}{2\kappa_1^2}d_1^T(M_1^{-1}-\zeta_1M_1^{-1}C_1)d_1\right]\times\zeta_1^{\alpha_{\zeta 1}-1}(1-\zeta_1)^{\beta_{\zeta 1}-1}
\end{aligned}$$

Let $\lambda_{1(j)}$ be the ordered eigenvalue of C_1, then

$$|I-\zeta_1C_1|=\prod_{j=1}^{n_1}(1-\zeta_1\lambda_{1(j)})$$

and

$$\log P(\zeta_1|rest)=C_0+\frac{1}{2}\sum_{j=1}^{n_1}\log(1-\zeta_1\lambda_{1(j)})+\frac{\zeta_1}{2\kappa_1^2}d_1^T(M_1^{-1}C_1)d_1+(\alpha_{\zeta 1}-1)\log\zeta_1+(\beta_{\zeta 1}-1)\log(1-\zeta_1)$$

Conditional distribution of κ_2^2

$$\begin{aligned}
P(\kappa_2^2|\text{rest}) &\propto p(d_2|\zeta_2,\kappa_2^2)\times p(\kappa_2^2)\\
&\propto \text{N}[d_2|\bar{0},\kappa_2^2(I-\zeta_2C_2)^{-1}M_2]\times \text{Inv-}\chi^2(\nu_{\kappa 2},S_{\kappa 2}^2)\\
&\propto (\kappa_2^2)^{-n_2/2}\exp\left[\tfrac{-1}{2\kappa_2^2}d_2^T(M_2^{-1}-\zeta_2M_2^{-1}C_2)d_2\right](\kappa_2^2)^{-(\nu_{\kappa 2}/2+1)}\exp\left(-\tfrac{\nu_{\kappa 2}S_{\kappa 2}^2}{2\kappa_2^2}\right)\\
&\propto (\kappa_2^2)^{-\left(\frac{n_2\nu_{\kappa 2}}{2}+1\right)}\exp\left\{\tfrac{-1}{2\kappa_2^2}[\nu_{\kappa 2}S_{\kappa 2}^2+d_2^T(M_2^{-1}-\zeta_2M_2^{-1}C_2)d_2]\right\}\\
&\propto \text{Inv-}\chi^2\{n_2+\nu_{\kappa 2},(n_2+\nu_{\kappa 2})^{-1}[\nu_{\kappa 2}S_{\kappa 2}^2+d_2^T(M_2^{-1}-\zeta_2M_2^{-1}C_2)d_2]\}
\end{aligned}$$

Conditional distribution of ζ_2

$$\begin{aligned} P(\zeta_2|\text{rest}) &\propto p(d_2|\zeta_2, \kappa_2^2) \times p(\zeta_2) \\ &\propto \mathrm{N}[d_2|\bar{0}, \kappa_2^2(I-\zeta_2 C_2)^{-1}M_2] \times \mathrm{beta}(\zeta_2|\alpha_{\zeta 2}, \beta_{\zeta 2}) \\ &\propto |I-\zeta_2 C_2|^{1/2} \exp\left[\tfrac{-1}{2\kappa_2^2} d_2^T (M_2^{-1} - \zeta_2 M_2^{-1} C_2) d_2\right] \times \zeta_2^{\alpha_{\zeta 2}-1}(1-\zeta_2)^{\beta_{\zeta 2}-1} \end{aligned}$$

Let $\lambda_{2(j)}$ be the ordered eigenvalue of C_2, then

$$|I-\zeta_2 C_2| = \prod_{j=1}^{n_2}(1-\zeta_2\lambda_{2(j)})$$

and

$$\log p(\zeta_2|rest) = C_0 + \frac{1}{2}\sum_{j=1}^{n_2}\log(1-\zeta_2\lambda_{2(j)}) + \frac{\zeta_2}{2\kappa_2^2} d_2^T(M_2^{-1}C_2)d_2 + (\alpha_{\zeta 2}-1)\log\zeta_2 + (\beta_{\zeta 2}-1)\log(1-\zeta_2)$$

And finally, the conditional distribution of α

$$\begin{aligned} P(\alpha|\text{rest}) &\propto p(S|d_1, d_2, \eta, \alpha, N) \times p(\alpha|\alpha_0, \beta_0) \\ &\propto \textstyle\prod_{i=1}^J \prod_{t=1}^T \mathrm{gamma}\left[S_{it}|N_{it}\exp(\lambda_{it}), \tfrac{N_{it}}{\alpha}\exp(2\lambda_{it})\right] \times \mathrm{gamma}(\alpha_0, \beta_0) \\ &\propto \textstyle\prod_{i=1}^J \prod_{t=1}^T \tfrac{1}{\gamma(\alpha N_{it})}[\alpha\exp(-\lambda_{it})]^{\alpha N_{it}} S_{it}^{(\alpha N_{it})-1} \exp[-\alpha\exp(-\lambda_{it})S_{it}] \\ &\quad \times \tfrac{\beta_0^{\alpha_0}}{\alpha_0}\alpha^{\alpha_0-1}\exp(-\beta_0\alpha) \\ \log(\alpha|\text{rest}) &= \textstyle\sum_{i=1}^J \sum_{t=1}^T \{-\log\mathrm{gamma}(\alpha N_{it}) + (\alpha N_{it})\log[\alpha\exp(-\lambda_{it})] \\ &\quad +(\alpha N_{it}-1)\log S_{it} - \alpha\exp(-\lambda_{it})S_{it}\} \\ &\quad +(\alpha_0-1)\log\alpha - \beta_0\alpha \end{aligned}$$

where

$$\lambda_{it} = x_i^T\eta + d_{1,t} + d_{2,i}$$

D. Posterior mean with upper and lower multiplication factors

D.1. Postal categories

Table D.1: Claim frequency for the postal categories with posterior mean, upper and lower multiplication factors and 95% posterior interval for the categories.

Claim frequency						
Nr. code	Region	Mean	L.b.m.f	U.b.m.f	95% lower	95% upper
1	Reykjavík	0,0743	0,8258	1,1868	0,0613	0,0882
2	Reykjavík Urban Region	0.0663	0.8435	1.1635	0.0560	0.0772
3	Large Urban Region	0.0644	0.8770	1.1205	0.0564	0.0721
4	Small Urban Region	0.0513	0.9156	1.0829	0.0470	0.0555
5	Rural near highway	0.0460	0.8658	1.1424	0.0398	0.0525
6	Rural area	0,0505	0.8019	1.2289	0.0405	0.0620
7	Banks and government	0.1101	0.6460	1.4450	0.0711	0.1591
8	Garbage	0.0751	0.6898	1.3778	0.0518	0.1034

Table D.2: Total claim size for the postal categories with posterior mean, upper and lower multiplication factors and 95% posterior interval for the categories.

Total claim size						
Nr. code	Region	Mean	L.b.m.f	U.b.m.f	95% lower	95% upper
1	Reykjavík	82291	0,9579	1,0460	78824	86074
2	Reykjavík Urban Region	82001	0.9426	1.0617	77297	87062
3	Large Urban Region	86079	0.9329	1.0719	80306	92264
4	Small Urban Region	88740	0.9662	1.0359	85740	91925
5	Rural near highway	86080	0.9500	1.0549	81778	90803
6	Rural area	86729	0.9401	1.0636	81532	92244
7	Banks and government	93742	0.9152	1.0937	85795	102525
8	Other	88912	0.9147	1.0951	81325	97367

Table D.3: Expected compensation for the postal categories with posterior mean, upper and lower multiplication factors and 95% posterior interval for the categories.

Expected compensation						
Nr. code	Region	Mean	L.b.m.f	U.b.m.f	95% lower	95% upper
1	Reykjavík	6113	0,7910	1,2414	4835	7589
2	Reykjavík Urban Region	5440	0.7951	1,2353	4325	6720
3	Large Urban Region	5540	0.8182	1,2010	4533	6654
4	Small Urban Region	4551	0.8846	1,1218	4026	5105
5	Rural near highway	3956	0.8225	1,2051	3253	4767
6	Rural area	4377	0,7538	1.3071	3299	5721
7	Banks and government	10321	0.5912	1,5803	6102	16310
8	Other	6674	0,6309	1.5089	4211	10071

D.2. Claim frequency

Table D.4: Table for expected claim frequency per number of policy years (Poisson) with posterior mean and upper and lower multiplication factor for 95% posterior interval. It also contains the numeric code which corresponds to the postal categories.

Claim frequency					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Reykjavík	101	0.0761	0.7260	1.3402	1
Reykjavík	103	0.0713	0.6781	1.4167	1
Reykjavík	104	0.0771	0.7232	1.3470	1
Reykjavík	105	0.0696	0.7279	1.3321	1
Reykjavík	107	0.0681	0.7109	1.3594	1
Reykjavík	108	0.0702	0.7288	1.3342	1
Reykjavík	109	0.0711	0.7262	1.3412	1
Reykjavík	110	0.0718	0.7247	1.3425	1
Reykjavík	111	0.0821	0.7227	1.3516	1
Reykjavík	112	0.0622	0.7258	1.3321	1
Reykjavík	113	0.0678	0.7023	1.3742	1
Reykjavík	116	0.1040	0.6347	1.5086	1
Reykjavík	150	0.1146	0.6357	1.4654	7
Reykjavík	155	0.1056	0.6357	1.4654	7
Seltjarnarnes	170	0.0753	0.6972	1.3888	2
Vogar	190	0.0593	0.7041	1.3923	4
Kópavogur	200	0.0652	0.7537	1.3017	2
Kópavogur	201	0.0607	0.7399	1.3174	2
Kópavogur	203	0.0632	0.7169	1.3559	2
Garðabær	210	0.0624	0.7414	1.3220	2
Hafnarfjörður	220	0.0716	0.7530	1.3026	2
Hafnarfjörður	221	0.0619	0.7372	1.3209	2
Álftanes	225	0.0716	0.6886	1.4019	2
Reykjanesbær	230	0.0717	0.7689	1.2812	3
Reykjanesbær	233	0.1185	0.6647	1.4646	3
Reykjanesbær	235	0.0545	0.6977	1.3645	3
Grindavík	240	0.0569	0.7243	1.3589	4
Sandgerði	245	0.0548	0.7201	1.3568	4
Garður	250	0.0568	0.6995	1.3938	4
Reykjanesbær	260	0.0625	0.7550	1.2991	3
Mosfellsbær	270	0.0652	0.7391	1.3207	2

Continued on next page

Table D.4 – continued from previous page

Claim frequency					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Akranes	300	0.0641	0.7625	1.2913	3
Akranes	301	0.0556	0.6615	1.4721	5
Borgarnes	310	0.0489	0.7202	1.3411	3
Borgarnes	311	0.0454	0.7306	1.3596	5
Reykholt	320	0.0544	0.6468	1.5038	6
Stykkishólmur	340	0.0508	0.7000	1.3986	4
Flatey á Breiðarfirði	345	0.0728	0.2879	2.5275	6
Grundarfjörður	350	0.0459	0.6657	1.4274	4
Ólafsvík	355	0.0553	0.6708	1.4531	4
Snæfellsbær	356	0.0760	0.5611	1.5988	4
Hellissandur	360	0.0518	0.6077	1.5297	4
Búðardalur	370	0.0459	0.6416	1.4340	4
Búðardalur	371	0.0395	0.6259	1.4873	6
Reykhólahreppur	380	0.0424	0.6224	1.5148	6
Ísafjörður	400	0.0423	0.7651	1.2946	4
Ísafjörður	401	0.0527	0.6712	1.4453	6
Hnífsdalur	410	0.0442	0.6688	1.4311	6
Bolungarvík	415	0.0450	0.6881	1.4102	4
Súðavík	420	0.0539	0.6903	1.3886	4
Flateyri	425	0.0496	0.6956	1.3699	4
Suðureyri	430	0.0492	0.6999	1.3639	4
Patreksfjörður	450	0.0422	0.7188	1.3580	4
Patreksfjörður	451	0.0476	0.6372	1.4619	6
Tálknafjörður	460	0.0364	0.4965	1.6817	4
Bíldudalur	465	0.0545	0.6743	1.4028	4
Þingeyri	470	0.0416	0.6992	1.3626	4
Þingeyri	471	0.0488	0.6623	1.4482	6
Staður	500	0.0499	0.6081	1.5253	5
Hólmavík	510	0.0478	0.6599	1.4613	4
Hólmavík	512	0.0708	0.5508	1.6742	6
Drangsnes	520	0.0666	0.5770	1.6381	4
	522	0.0302	0.6422	1.4867	6
	523	0.0636	0.6301	1.4958	6
Árneshreppur	524	0.0842	0.6214	1.5233	6
Hvammstangi	530	0.0534	0.6848	1.4307	4
Hvammstangi	531	0.0427	0.6462	1.4628	5
Blönduós	540	0.0455	0.6618	1.4497	4

Continued on next page

Table D.4 – continued from previous page

Claim frequency					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Blönduós	541	0.0442	0.6081	1.5468	5
Skagaströnd	545	0.0461	0.5134	1.7170	4
Sauðárkrókur	550	0.0540	0.7191	1.3692	4
Sauðárkrókur	551	0.0327	0.6349	1.4518	5
Varmahlíð	560	0.0424	0.6569	1.4447	4
Hofsós	565	0.0487	0.5813	1.5557	4
Hofsós	566	0.0642	0.5631	1.6279	5
Fljót	570	0.0552	0.5647	1.6126	6
Siglufjörður	580	0.0367	0.6190	1.5262	6
Akureyri	600	0.0533	0.7862	1.2580	3
Akureyri	601	0.0454	0.7350	1.3397	5
Akureyri	603	0.0607	0.7709	1.2764	3
Grenivík	610	0.0644	0.6103	1.5506	4
Grímsey	611	0.1183	0.3828	2.1529	4
Dalvík	620	0.0368	0.6948	1.3796	4
Dalvík	621	0.0307	0.5837	1.5229	5
Ólafsfjörður	625	0.0582	0.6686	1.4554	4
Hrísey	630	0.0630	0.3539	2.1087	4
Húsavík	640	0.0517	0.7322	1.3504	4
Húsavík	641	0.0488	0.6399	1.4953	6
Fosshóll	645	0.0559	0.6180	1.5301	5
Laugar	650	0.0568	0.5575	1.5766	4
Mývatn	660	0.0433	0.5912	1.5311	4
Kópasker	670	0.0465	0.5601	1.5794	4
Kópasker	671	0.0389	0.5497	1.6201	6
Raufarhöfn	675	0.0428	0.5685	1.5675	4
Þórshöfn	680	0.0400	0.5998	1.5232	4
Þórshöfn	681	0.0358	0.5540	1.6215	6
Bakkafjörður	685	0.0655	0.4834	1.8024	4
Vopnafjörður	690	0.0419	0.6751	1.4370	4
Egilsstaðir	700	0.0473	0.7494	1.3086	3
Egilsstaðir	701	0.0288	0.7030	1.3597	5
Seyðisfjörður	710	0.0395	0.6883	1.3938	4
Mjóafjörður	715	0.0642	0.6418	1.4811	6
Borgarfjörður eystri	720	0.0442	0.5327	1.6329	4
Reyðarfjörður	730	0.0429	0.7127	1.3813	4
Eskifjörður	735	0.0464	0.7147	1.3910	4

Continued on next page

Table D.4 – continued from previous page

Claim frequency					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Neskaupsstaður	740	0.0418	0.7078	1.3776	4
Fáskrúðsfjörður	750	0.0378	0.6816	1.3938	4
Stöðvarfjörður	755	0.0398	0.6593	1.4109	4
Breiðdalsvík	760	0.0564	0.6786	1.4326	4
Djúpivogur	765	0.0397	0.7017	1.3670	4
Höfn	780	0.0502	0.7157	1.3667	4
Höfn	781	0.0323	0.5776	1.5598	5
Öræfi	785	0.0377	0.5147	1.6389	5
Selfoss	800	0.0635	0.7552	1.3014	3
Selfoss	801	0.0483	0.7269	1.3654	5
Hveragerði	810	0.0629	0.7133	1.3665	3
Þorlákshöfn	815	0.0515	0.7028	1.3854	4
Eyrarbakki	820	0.0638	0.6805	1.4378	4
Stokkseyri	825	0.0579	0.6832	1.4317	4
Laugarvatn	840	0.0528	0.4670	1.8157	5
Flúðir	845	0.0509	0.6120	1.5343	4
Hella	850	0.0489	0.7008	1.3710	4
Hella	851	0.0395	0.6839	1.3921	5
Hvolsvöllur	860	0.0471	0.6859	1.4090	4
Hvolsvöllur	861	0.0395	0.6690	1.4309	5
Vík	870	0.0647	0.6149	1.5255	4
Vík	871	0.0548	0.6153	1.5479	5
Kirkjubæjarklaustur	880	0.0516	0.6814	1.4243	4
Vestmannaeyjar	900	0.0491	0.6958	1.3839	4
	999	0.0751	0.6760	1.3970	8

D.3. Total claim size

Table D.5: Table for expected claim size per number of policy years (gamma) with posterior mean and upper and lower multiplication factor for 95% posterior interval. It also contains the numeric code which corresponds to the postal categories.

Claim size					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Reykjavík	101	82291	0.8507	1.1898	1
Reykjavík	103	82291	0.8507	1.1898	1
Reykjavík	104	82291	0.8507	1.1898	1
Reykjavík	105	82291	0.8507	1.1898	1
Reykjavík	107	82291	0.8507	1.1898	1
Reykjavík	108	82291	0.8507	1.1898	1
Reykjavík	109	82291	0.8507	1.1898	1
Reykjavík	110	82291	0.8507	1.1898	1
Reykjavík	111	82291	0.8507	1.1898	1
Reykjavík	112	82291	0.8507	1.1898	1
Reykjavík	113	82291	0.8507	1.1898	1
Reykjavík	116	82291	0.8507	1.1898	1
Reykjavík	150	93742	0.8337	1.2026	7
Reykjavík	155	93742	0.8337	1.2026	7
Seltjarnarnes	170	82001	0.8468	1.1944	2
Vogar	190	88740	0.8531	1.1847	4
Kópavogur	200	82001	0.8468	1.1944	2
Kópavogur	201	82001	0.8468	1.1944	2
Kópavogur	203	82001	0.8468	1.1944	2
Garðabær	210	82001	0.8468	1.1944	2
Hafnarfjörður	220	82001	0.8468	1.1944	2
Hafnarfjörður	221	82001	0.8468	1.1944	2
Álftanes	225	82001	0.8468	1.1944	2
Reykjanesbær	230	86079	0.8407	1.2004	3
Reykjanesbær	233	86079	0.8407	1.2004	3
Reykjanesbær	235	86079	0.8407	1.2004	3
Grindavík	240	88740	0.8531	1.1847	4
Sandgerði	245	88740	0.8531	1.1847	4
Garður	250	88740	0.8531	1.1847	4
Reykjanesbær	260	86079	0.8407	1.2004	3
Mosfellsbær	270	82001	0.8468	1.1944	2
Akranes	300	86079	0.8407	1.2004	3
Akranes	301	86080	0.8484	1.1912	5
Borgarnes	310	86079	0.8407	1.2004	3

Continued on next page

Table D.5 – continued from previous page

Claim size					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Borgarnes	311	86080	0.8484	1.1912	5
Reykholt	320	86729	0.8457	1.1950	6
Stykkishólmur	340	88740	0.8531	1.1847	4
Flatey á Breiðarfirði	345	86080	0.8484	1.1912	6
Grundarfjörður	350	88740	0.8531	1.1847	4
Ólafsvík	355	88740	0.8531	1.1847	4
Snæfellsbær	356	88740	0.8531	1.1847	4
Hellissandur	360	88740	0.8531	1.1847	4
Búðardalur	370	88740	0.8531	1.1847	4
Búðardalur	371	86729	0.8457	1.1950	6
Reykhólahreppur	380	86729	0.8457	1.1950	6
Ísafjörður	400	88740	0.8531	1.1847	4
Ísafjörður	401	86729	0.8457	1.1950	6
Hnífsdalur	410	86729	0.8457	1.1950	6
Bolungarvík	415	88740	0.8531	1.1847	4
Súðavík	420	88740	0.8531	1.1847	4
Flateyri	425	88740	0.8531	1.1847	4
Suðureyri	430	88740	0.8531	1.1847	4
Patreksfjörður	450	88740	0.8531	1.1847	4
Patreksfjörður	451	86729	0.8457	1.1950	6
Tálknafjörður	460	88740	0.8531	1.1847	4
Bíldudalur	465	88740	0.8531	1.1847	4
Þingeyri	470	88740	0.8531	1.1847	4
Þingeyri	471	86729	0.8457	1.1950	6
Staður	500	86080	0.8484	1.1912	5
Hólmavík	510	88740	0.8531	1.1847	4
Hólmavík	512	86729	0.8457	1.1950	6
Drangsnes	520	88740	0.8531	1.1847	4
	522	86729	0.8457	1.1950	6
	523	86729	0.8457	1.1950	6
Árneshreppur	524	86729	0.8457	1.1950	6
Hvammstangi	530	88740	0.8531	1.1847	4
Hvammstangi	531	86080	0.8484	1.1912	5
Blönduós	540	88740	0.8531	1.1847	4
Blönduós	541	86080	0.8484	1.1912	5
Skagaströnd	545	88740	0.8531	1.1847	4
Sauðárkrókur	550	88740	0.8531	1.1847	4

Continued on next page

Table D.5 – continued from previous page

Claim size					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Sauðárkrókur	551	86080	0.8484	1.1912	5
Varmahlíð	560	88740	0.8531	1.1847	4
Hofsós	565	88740	0.8531	1.1847	4
Hofsós	566	86080	0.8484	1.1912	5
Fljót	570	86729	0.8457	1.1950	6
Siglufjörður	580	86729	0.8457	1.1950	6
Akureyri	600	86079	0.8407	1.2004	3
Akureyri	601	86080	0.8484	1.1912	5
Akureyri	603	86079	0.8407	1.2004	3
Grenivík	610	88740	0.8531	1.1847	4
Grímsey	611	88740	0.8531	1.1847	4
Dalvík	620	88740	0.8531	1.1847	4
Dalvík	621	86080	0.8484	1.1912	5
Ólafsfjörður	625	88740	0.8531	1.1847	4
Hrísey	630	88740	0.8531	1.1847	4
Húsavík	640	88740	0.8531	1.1847	4
Húsavík	641	86729	0.8457	1.1950	6
Fosshóll	645	86080	0.8484	1.1912	5
Laugar	650	88740	0.8531	1.1847	4
Mývatn	660	88740	0.8531	1.1847	4
Kópasker	670	88740	0.8531	1.1847	4
Kópasker	671	86729	0.8457	1.1950	6
Raufarhöfn	675	88740	0.8531	1.1847	4
Þórshöfn	680	88740	0.8531	1.1847	4
Þórshöfn	681	86729	0.8457	1.1950	6
Bakkafjörður	685	88740	0.8531	1.1847	4
Vopnafjörður	690	88740	0.8531	1.1847	4
Egilsstaðir	700	86079	0.8407	1.2004	3
Egilsstaðir	701	86080	0.8484	1.1912	5
Seyðisfjörður	710	88740	0.8531	1.1847	4
Mjóafjörður	715	86729	0.8457	1.1950	6
Borgarfjörður eystri	720	88740	0.8531	1.1847	4
Reyðarfjörður	730	88740	0.8531	1.1847	4
Eskifjörður	735	88740	0.8531	1.1847	4
Neskaupsstaður	740	88740	0.8531	1.1847	4
Fáskrúðsfjörður	750	88740	0.8531	1.1847	4
Stöðvarfjörður	755	88740	0.8531	1.1847	4

Continued on next page

Table D.5 – continued from previous page

Claim size					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Breiðdalsvík	760	88740	0.8531	1.1847	4
Djúpivogur	765	88740	0.8531	1.1847	4
Höfn	780	88740	0.8531	1.1847	4
Höfn	781	86080	0.8484	1.1912	5
Öræfi	785	86080	0.8484	1.1912	5
Selfoss	800	86079	0.8407	1.2004	3
Selfoss	801	86080	0.8484	1.1912	5
Hveragerði	810	86079	0.8407	1.2004	3
Þorlákshöfn	815	88740	0.8531	1.1847	4
Eyrarbakki	820	88740	0.8531	1.1847	4
Stokkseyri	825	88740	0.8531	1.1847	4
Laugarvatn	840	86080	0.8484	1.1912	5
Flúðir	845	88740	0.8531	1.1847	4
Hella	850	88740	0.8531	1.1847	4
Hella	851	86080	0.8484	1.1912	5
Hvolsvöllur	860	88740	0.8531	1.1847	4
Hvolsvöllur	861	86080	0.8484	1.1912	5
Vík	870	88740	0.8531	1.1847	4
Vík	871	86080	0.8484	1.1912	5
Kirkjubæjarklaustur	880	88740	0.8531	1.1847	4
Vestmannaeyjar	900	88740	0.8531	1.1847	4
	999	88912	0.8340	1.2054	8

D.4. Expected compensation

Table D.6: Table for expected value per number of policy years with posterior mean and upper and lower multiplication factor for 95% posterior interval. It also contains the numeric code which corresponds to the postal categories.

Expected value					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Reykjavík	101	6265	0.6965	1.3980	1

Continued on next page

Table D.6 – continued from previous page

Expected value					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Reykjavík	103	5871	0.6534	1.4641	1
Reykjavík	104	6345	0.6942	1.4029	1
Reykjavík	105	5725	0.6996	1.3951	1
Reykjavík	107	5606	0.6840	1.4177	1
Reykjavík	108	5776	0.6983	1.3932	1
Reykjavík	109	5848	0.6975	1.3986	1
Reykjavík	110	5912	0.6975	1.4018	1
Reykjavík	111	6753	0.6918	1.4092	1
Reykjavík	112	5117	0.6991	1.3915	1
Reykjavík	113	5583	0.6770	1.4289	1
Reykjavík	116	8558	0.6114	1.5495	1
Reykjavík	150	10743	0.6115	1.5180	7
Reykjavík	155	9898	0.6115	1.5180	7
Seltjarnarnes	170	6177	0.6711	1.4458	2
Vogar	190	5261	0.6775	1.4417	4
Kópavogur	200	5344	0.7170	1.3720	2
Kópavogur	201	4976	0.7069	1.3829	2
Kópavogur	203	5185	0.6862	1.4174	2
Garðabær	210	5120	0.7071	1.3876	2
Hafnarfjörður	220	5870	0.7174	1.3726	2
Hafnarfjörður	221	5075	0.7047	1.3869	2
Álftanes	225	5868	0.6610	1.4563	2
Reykjanesbær	230	6174	0.7249	1.3519	3
Reykjanesbær	233	10204	0.6396	1.5151	3
Reykjanesbær	235	4689	0.6683	1.4250	3
Grindavík	240	5049	0.6949	1.4120	4
Sandgerði	245	4865	0.6929	1.4092	4
Garður	250	5040	0.6750	1.4390	4
Reykjanesbær	260	5376	0.7169	1.3662	3
Mosfellsbær	270	5346	0.7054	1.3845	2
Akranes	300	5521	0.7198	1.3586	3
Akranes	301	4789	0.6366	1.5170	5
Borgarnes	310	4212	0.6871	1.4020	3
Borgarnes	311	3912	0.6972	1.4122	5
Reykholt	320	4717	0.6220	1.5421	6
Stykkishólmur	340	4509	0.6732	1.4492	4
Flatey á Breiðarfirði	345	6261	0.2817	2.5489	6

Continued on next page

Table D.6 – continued from previous page

Expected value					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Grundarfjörður	350	4072	0.6446	1.4695	4
Ólafsvík	355	4903	0.6478	1.4919	4
Snæfellsbær	356	6747	0.5470	1.6377	4
Hellissandur	360	4594	0.5894	1.5702	4
Búðardalur	370	4071	0.6211	1.4802	4
Búðardalur	371	3430	0.6054	1.5333	6
Reykhólahreppur	380	3675	0.6034	1.5618	6
Ísafjörður	400	3758	0.7286	1.3583	4
Ísafjörður	401	4567	0.6479	1.4931	6
Hnífsdalur	410	3834	0.6442	1.4791	6
Bolungarvík	415	3993	0.6644	1.4580	4
Súðavík	420	4780	0.6654	1.4388	4
Flateyri	425	4400	0.6715	1.4223	4
Suðureyri	430	4364	0.6765	1.4147	4
Patreksfjörður	450	3741	0.6890	1.4083	4
Patreksfjörður	451	4126	0.6151	1.5084	6
Tálknafjörður	460	3230	0.4872	1.7250	4
Bíldudalur	465	4839	0.6503	1.4478	4
Þingeyri	470	3688	0.6753	1.4164	4
Þingeyri	471	4232	0.6376	1.5000	6
Staður	500	4292	0.5880	1.5627	5
Hólmavík	510	4238	0.6357	1.5028	4
Hólmavík	512	6139	0.5379	1.7158	6
Drangsnes	520	5911	0.5607	1.6655	4
	522	2618	0.6184	1.5307	6
	523	5519	0.6095	1.5426	6
Árneshreppur	524	7299	0.6005	1.5673	6
Hvammstangi	530	4739	0.6583	1.4730	4
Hvammstangi	531	3672	0.6218	1.5102	5
Blönduós	540	4035	0.6402	1.4975	4
Blönduós	541	3804	0.5895	1.5918	5
Skagaströnd	545	4092	0.5021	1.7501	4
Sauðárkrókur	550	4794	0.6900	1.4224	4
Sauðárkrókur	551	2814	0.6121	1.5042	5
Varmahlíð	560	3760	0.6357	1.4934	4
Hofsós	565	4325	0.5652	1.5960	4
Hofsós	566	5523	0.5497	1.6655	5

Continued on next page

Table D.6 – continued from previous page

Expected value					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Fljót	570	4791	0.5479	1.6547	6
Siglufjörður	580	3183	0.5989	1.5714	6
Akureyri	600	4585	0.7408	1.3297	3
Akureyri	601	3908	0.6995	1.3963	5
Akureyri	603	5225	0.7283	1.3452	3
Grenivík	610	5715	0.5933	1.5864	4
Grímsey	611	10497	0.3754	2.1759	4
Dalvík	620	3264	0.6712	1.4256	4
Dalvík	621	2640	0.5674	1.5730	5
Ólafsfjörður	625	5165	0.6431	1.5023	4
Hrísey	630	5591	0.3487	2.1378	4
Húsavík	640	4587	0.7015	1.4031	4
Húsavík	641	4230	0.6201	1.5440	6
Fosshóll	645	4814	0.6010	1.5681	5
Laugar	650	5043	0.5446	1.6146	4
Mývatn	660	3843	0.5763	1.5665	4
Kópasker	670	4131	0.5453	1.6206	4
Kópasker	671	3371	0.5341	1.6593	6
Raufarhöfn	675	3797	0.5524	1.6094	4
Þórshöfn	680	3551	0.5820	1.5724	4
Þórshöfn	681	3109	0.5366	1.6647	6
Bakkafjörður	685	5813	0.4724	1.8329	4
Vopnafjörður	690	3719	0.6498	1.4844	4
Egilsstaðir	700	4068	0.7112	1.3732	3
Egilsstaðir	701	2479	0.6750	1.4179	5
Seyðisfjörður	710	3503	0.6659	1.4435	4
Mjóafjörður	715	5569	0.6181	1.5222	6
Borgarfjörður eystri	720	3920	0.5201	1.6651	4
Reyðarfjörður	730	3803	0.6864	1.4316	4
Eskifjörður	735	4120	0.6875	1.4353	4
Neskaupsstaður	740	3712	0.6819	1.4257	4
Fáskrúðsfjörður	750	3356	0.6583	1.4489	4
Stöðvarfjörður	755	3530	0.6375	1.4596	4
Breiðdalsvík	760	5008	0.6567	1.4775	4
Djúpivogur	765	3519	0.6764	1.4142	4
Höfn	780	4453	0.6872	1.4184	4
Höfn	781	2784	0.5612	1.6010	5

Continued on next page

Table D.6 – continued from previous page

Expected value					
Region	**Postal code**	**Mean**	**L.b. multiplication factor**	**U.b. multiplication factor**	**Numeric code**
Öræfi	785	3249	0.5029	1.6814	5
Selfoss	800	5469	0.7160	1.3673	3
Selfoss	801	4161	0.6941	1.4147	5
Hveragerði	810	5416	0.6813	1.4274	3
Þorlákshöfn	815	4572	0.6762	1.4303	4
Eyrarbakki	820	5664	0.6553	1.4879	4
Stokkseyri	825	5134	0.6573	1.4734	4
Laugarvatn	840	4542	0.4569	1.8458	5
Flúðir	845	4517	0.5932	1.5760	4
Hella	850	4339	0.6772	1.4176	4
Hella	851	3397	0.6595	1.4483	5
Hvolsvöllur	860	4184	0.6615	1.4503	4
Hvolsvöllur	861	3399	0.6442	1.4829	5
Vík	870	5737	0.5977	1.5686	4
Vík	871	4716	0.5973	1.5904	5
Kirkjubæjarklaustur	880	4583	0.6562	1.4721	4
Vestmannaeyjar	900	4354	0.6701	1.4302	4
	999	6674	0.6488	1.4547	8

E. Figures for claim frequency and expected compensation

E.1. Claim frequency

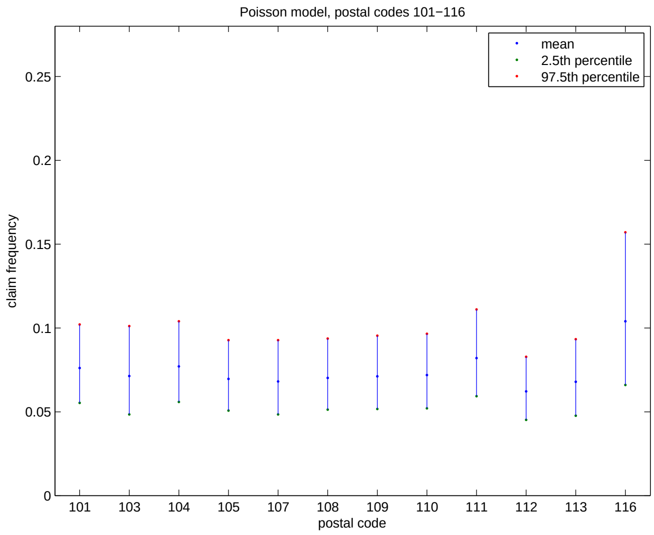

Figure E.1: Average claim frequency with 95% posterior interval for postal codes 101-116.

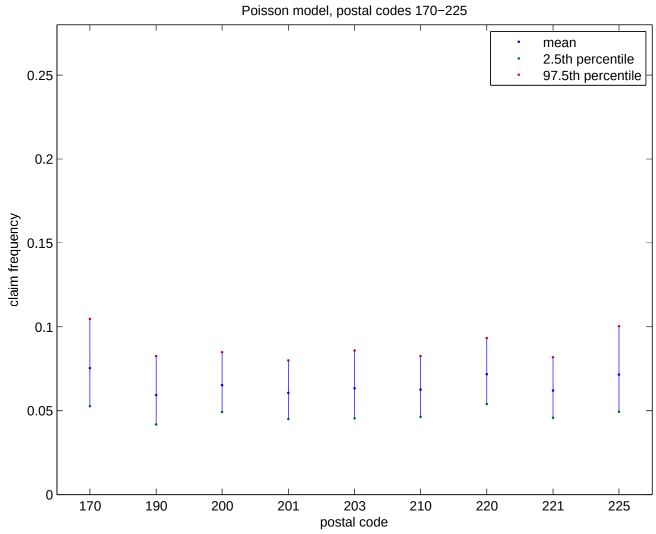

Figure E.2: Average claim frequency with 95% posterior interval for postal codes 170-225.

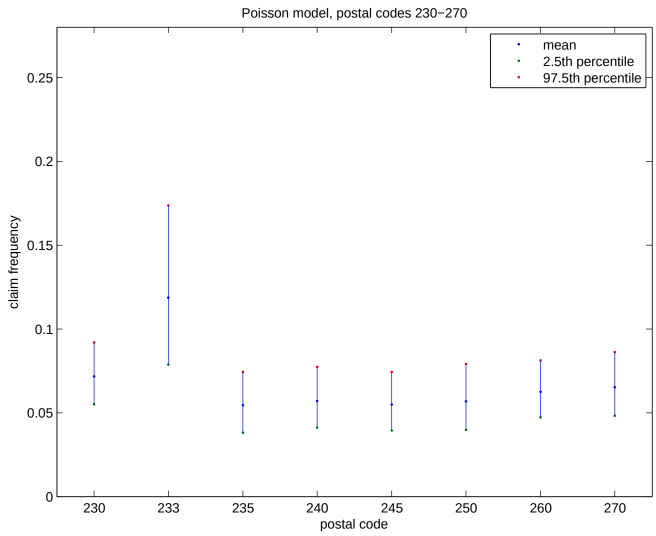

Figure E.3: Average claim frequency with 95% posterior interval for postal codes 230-270.

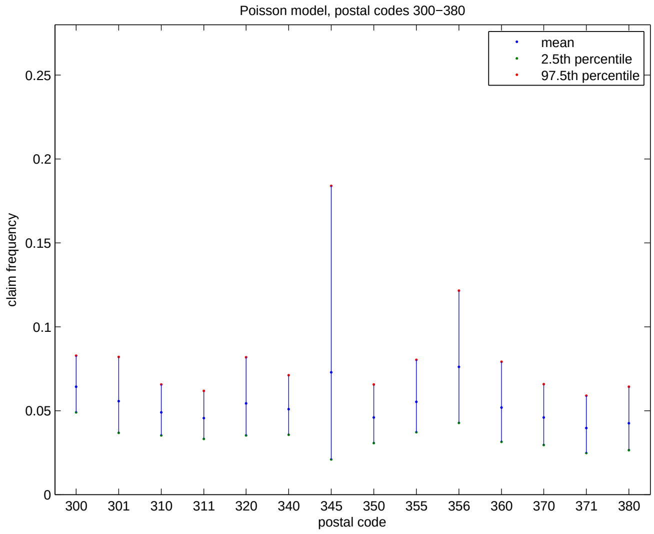

Figure E.4: Average claim frequency with 95% posterior interval for postal codes 300-380.

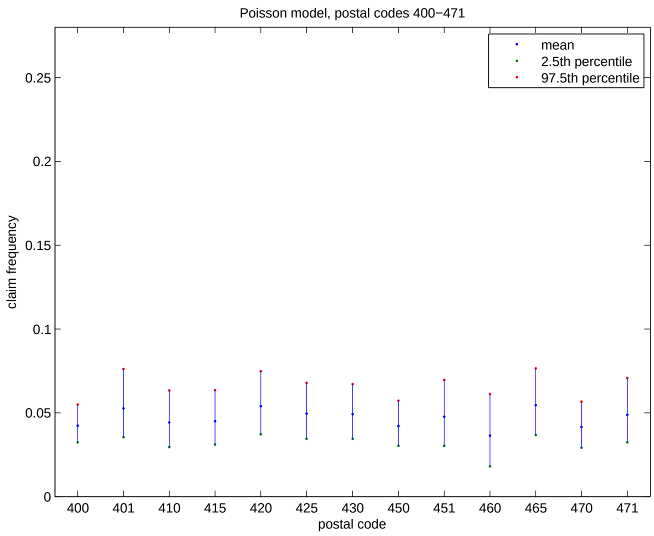

Figure E.5: Average claim frequency with 95% posterior interval for postal codes 400-471.

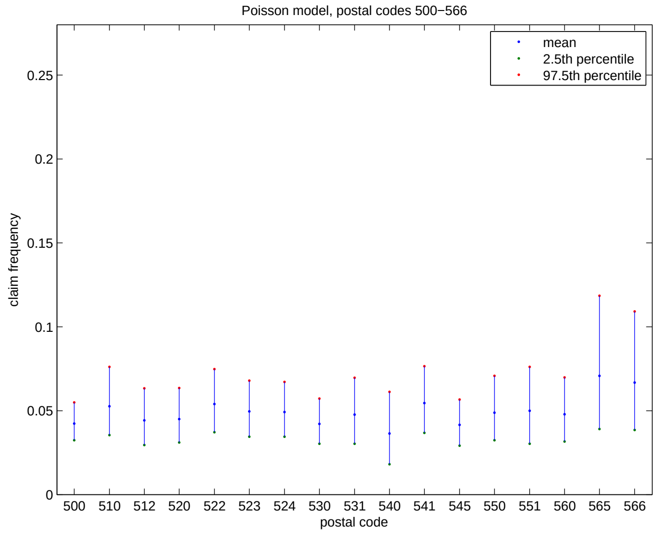

Figure E.6: Average claim frequency with 95% posterior interval for postal codes 500-566.

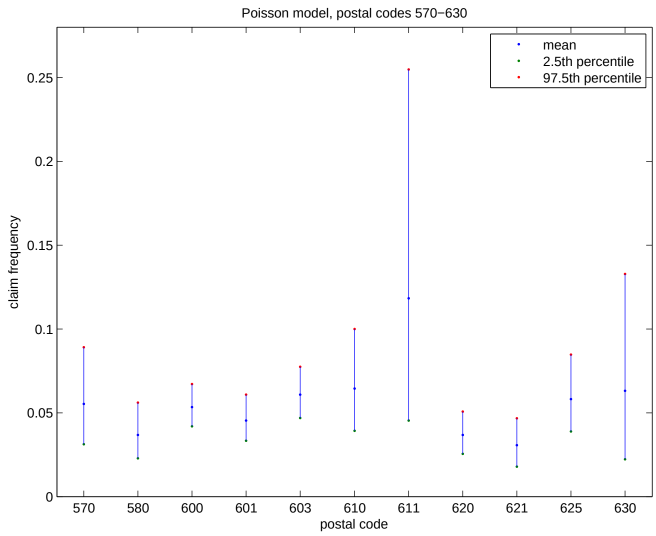

Figure E.7: Average claim frequency with 95% posterior interval for postal codes 570-630.

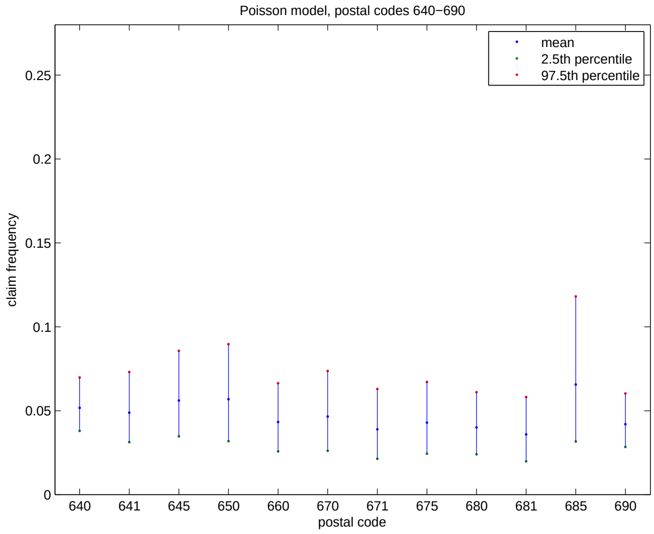

Figure E.8: Average claim frequency with 95% posterior interval for postal codes 640-690.

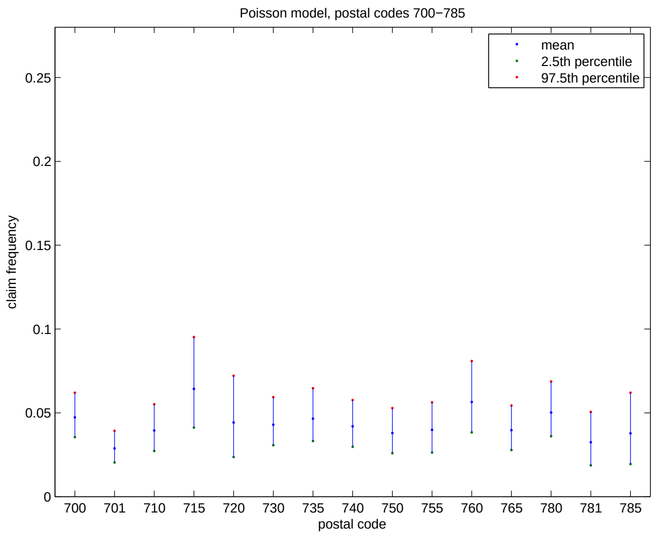

Figure E.9: Average claim frequency with 95% posterior interval for postal codes 700-785.

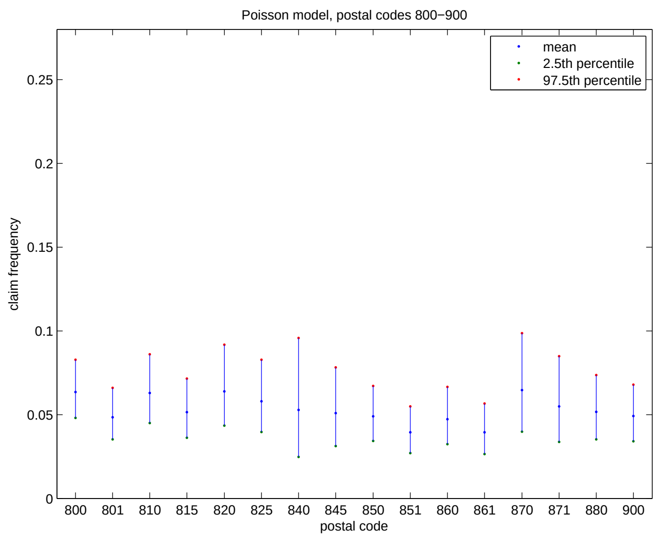

Figure E.10: Average claim frequency with 95% posterior interval for postal codes 800-900.

E.2. Expected compensation

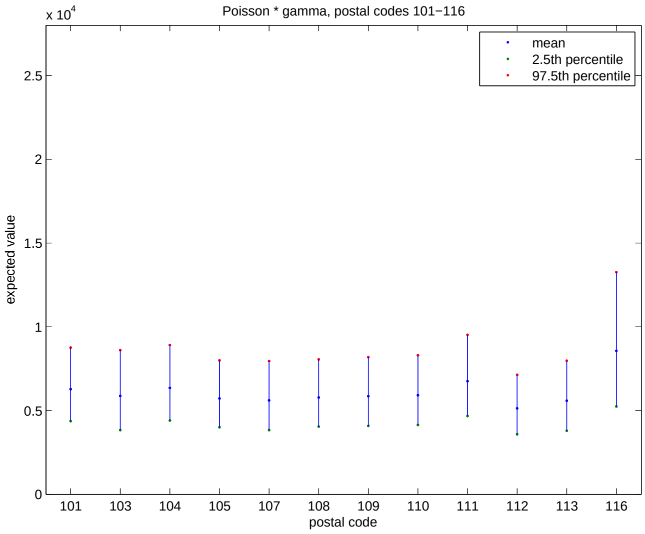

Figure E.11: Expected compensation with 95% posterior interval for postal codes 101-116.

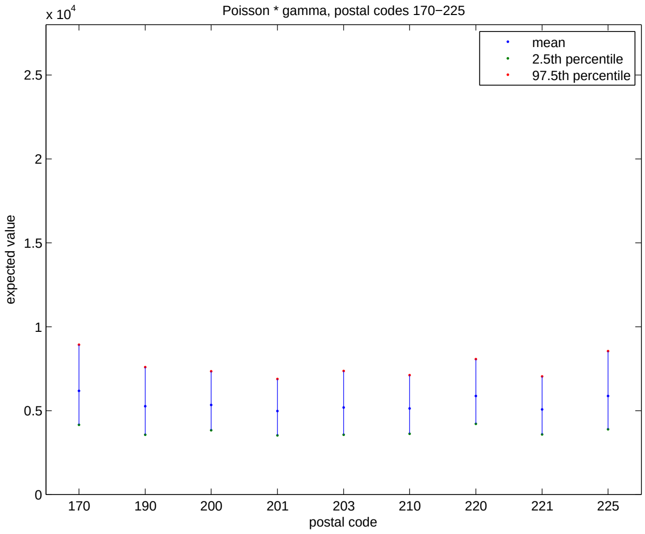

Figure E.12: Expected compensation with 95% posterior interval for postal codes 170-225.

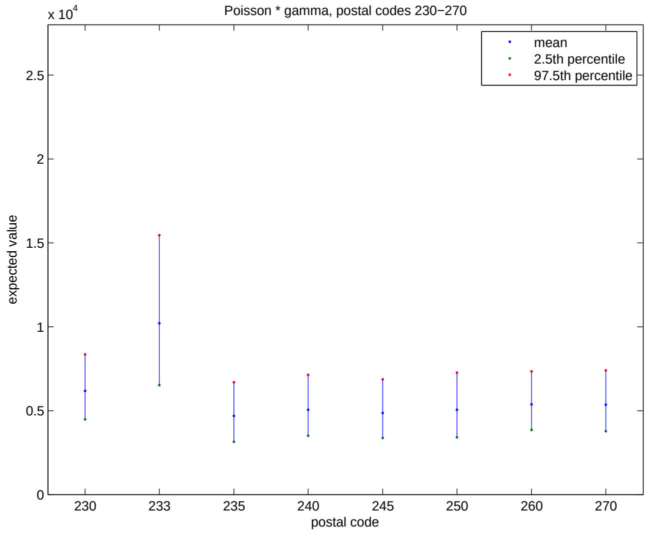

Figure E.13: Expected compensation with 95% posterior interval for postal codes 230-270.

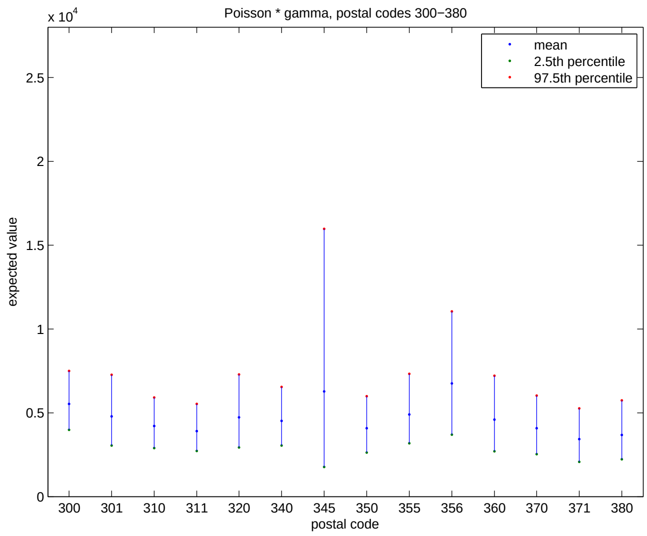

Figure E.14: Expected compensation with 95% posterior interval for postal codes 300-380.

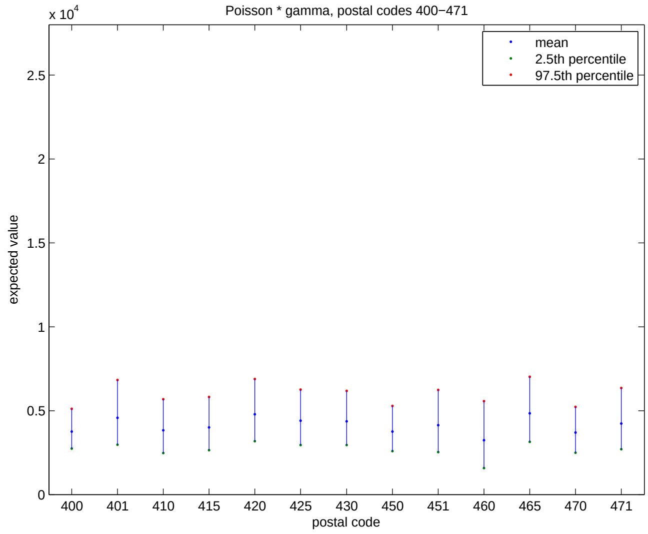

Figure E.15: Expected compensation with 95% posterior interval for postal codes 400-471.

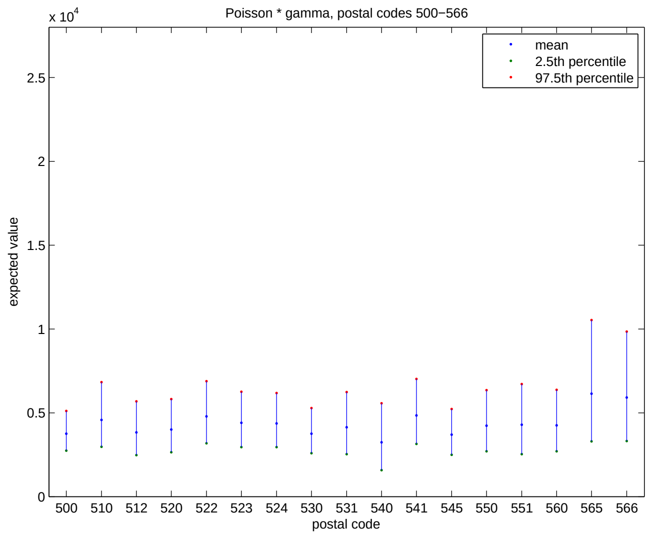

Figure E.16: Expected compensation with 95% posterior interval for postal codes 500-566.

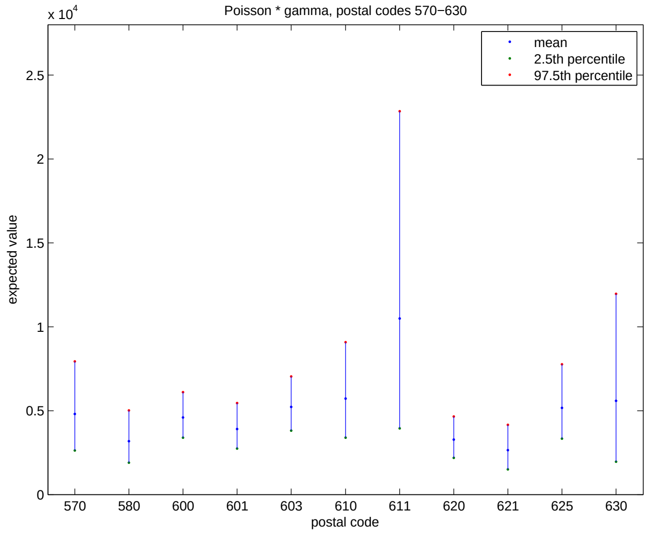

Figure E.17: Expected compensation with 95% posterior interval for postal codes 570-630.

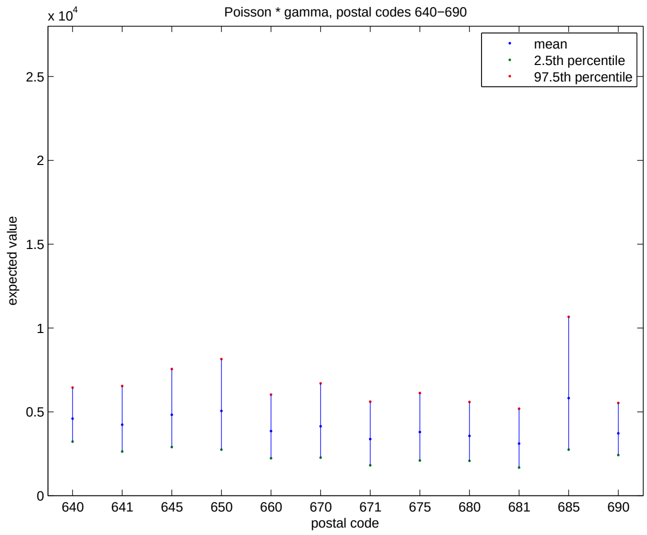

Figure E.18: Expected compensation with 95% posterior interval for postal codes 640-690.

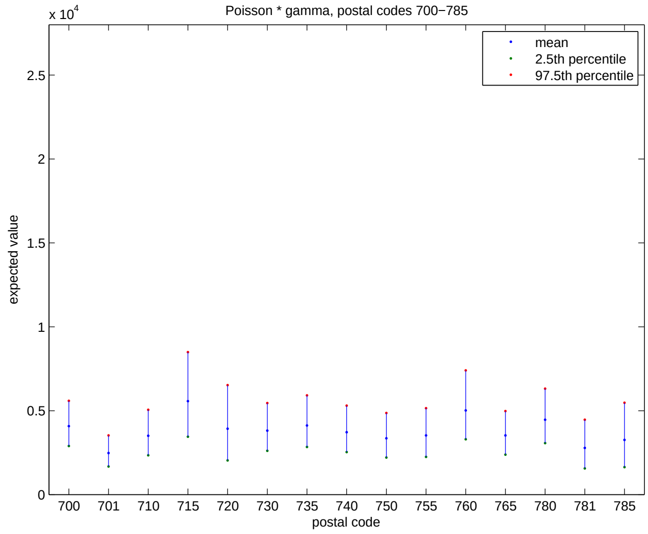

Figure E.19: Expected compensation with 95% posterior interval for postal codes 700-785.

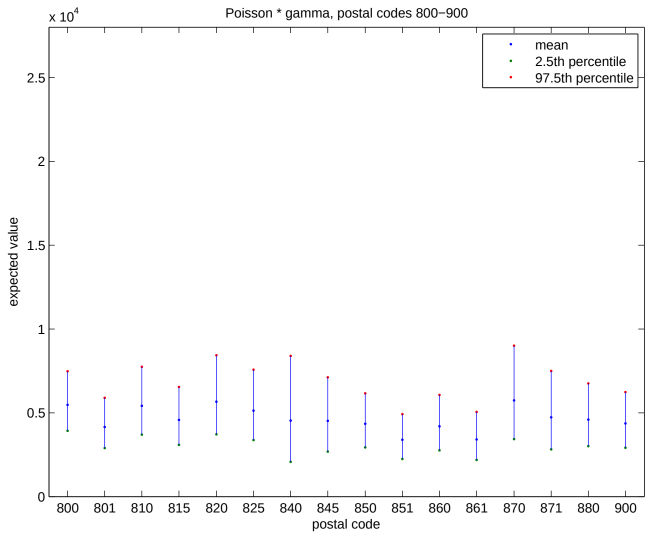

Figure E.20: Expected compensation with 95% posterior interval for postal codes 800-900.