Bayesian Quantile Analysis for Loss Reserving Models

Author: Fábio Vitor Generoso Vieira

Supervisor: Carlos Antonio Abanto-Valle

Department of Statistical Methods
Institute of Mathematics
Federal University of Rio de Janeiro

Rio de Janeiro - RJ
June 2019

Bayesian Quantile Analysis for Loss Reserving Models

Fábio Vitor Generoso Vieira

Master dissertation submitted to the Institute of Mathematics of the Federal University of Rio de Janeiro as part of the requisites to obtain the Statistics master's degree.

Supervisor: Carlos A. Abanto-Valle

Rio de Janeiro, RJ - Brazil

June 2019

Bayesian Quantile Analysis for Loss Reserving Models

Fábio Vitor Generoso Vieira

Master dissertation submitted to the Institute of Mathematics of the Federal University of Rio de Janeiro as part of the requisites to obtain the Statistics master's degree.

Approved by:

President - Prof. Carlos A. Abanto-Valle
Federal University of Rio de Janeiro - PhD

Prof. Mariane Branco Alves
Federal University of Rio de Janeiro - PhD

Prof. Víctor Hugo Lachos Dávila
University of Connecticut - PhD

Rio de Janeiro, RJ - Brazil

June 2019

CIP - Catalogação na Publicação

V657b

Vieira, Fabio Vitor Generoso
Bayesian quantile analysis for loss reserving models / Fabio Vitor Generoso Vieira. -- Rio de Janeiro, 2019.
73 f.

Orientador: Carlos Antonio Abanto-Valle.
Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Instituto de Matemática, Programa de Pós-Graduação em Estatística, 2019.

1. Loss reserving. 2. Bayesian inference. 3. Quantile analysis. 4. Risk management. I. Abanto Valle, Carlos Antonio, orient. II. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

Federal University of Rio de Janeiro

Bayesian Quantile Analysis for Loss Reserving Models

Fábio Vitor Generoso Vieira

2019

Acknowledgments

First and foremost, I would like to take this opportunity to express my gratitude to the Institute of Mathematics of the Federal University of Rio de Janeiro for making it possible for me to take my first step towards an academic career.

I must thank my supervisor, Professor Carlos Abanto-Valle, for his patience and support. He has provided me academic guidance that will shape the type of researcher I am going to be. The benefits taken from our interactions will keep being the building blocks of my personality as an academic.

Also, I dearly acknowledge the role of my family during my studies. My mother and my two sisters were the ones that kept me standing on solid ground when the hardest times tried to overthrow me.

Finally, I would like to thank the Brazilian government for funding my research through the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (Capes). Without this financial support my work would have never been completed and my goal of obtaining a Master's degree would be considerably more difficult to achieve.

Abstract

Bayesian Quantile Analysis for Loss Reserving Models

Insurance companies rely heavily on historical data to forecast appropriate levels of financial reserves to pay outstanding claims and settlement related costs. These data usually contain irregular and extreme values. Existing models focus on the computation of conditional means as loss reserving estimates. Such measures suffer from a lack of robustness in the presence of outliers. In this dissertation, a class of loss reserving models based on the quantile regression approach is developed with the main purpose of modeling the tails of the claims distribution. The basic idea is to allow the random components to follow asymmetric Laplace distributions. This design allows the prediction of not only central reserves estimates, such as the median, but also the projection of a region where future observations are more likely to be, by forecasting extreme conditional quantiles. The Bayesian paradigm using Markov chain Monte Carlo algorithms was adopted to compensate for uncertainty in the parameter estimation process. This approach has also the advantage to easily produce predictive distributions that give a better perception of the distribution tail behavior which could eventually help practitioners to understand their risk exposure and assist in risk transferring activities. Applications with real-world data were made to illustrate the methodology.

Keywords: Asymmetric Laplace distribution, loss reserving models, insurance risk management, Bayesian inference.

Resumo

Análise Quantílica Bayesiana para Modelos de Provisão de Sinistros

Companhias de seguros dependem em grande parte de observações passadas para prever níveis apropriados de reservas para pagar indenizações e demais custos associados à liquidação de sinistros. Usualmente, esses dados contêm valores irregulares e extremos. Os modelos existentes focam no cálculo de esperanças condicionais como estimadores para as reservas. No entanto, essas medidas carecem de robustez na presença de *outilers*. Nesta dissertação uma classe de modelos de provisão de sinistros é desenvolvida com base no método de regressão quantílica. O proposito principal é modelar as caudas da distribuição de indenizações. A ideia básica é permitir que os resíduos assumam uma distribuição de Laplace assimétrica. Esse desenho permite não só a previsão de valores no centro da distribuição, como a mediana, mas também a projeção de uma região onde as observações futuras são mais prováveis, através da previsão de quantis extremos. O paradigma Bayesiano , usando métodos de Monte Carlo via cadeias de Markov, foi adotado com o objetivo de compensar a incerteza no processo de estimação. Essa abordagem tem a vantagem de produzir distribuições preditivas com relativa facilidade, as quais dão uma melhor percepção do comportamento das caudas da distribuição. Esse método pode ser usado para compreender a exposição de uma carteira ao risco, além de possivelmente auxiliar em tarefas de transferência de risco. Aplicações com dados reais foram feitas para ilustrar a metodologia.

Palavras chave: Distribuição de Laplace assimétrica, modelos de provisão de sinistros, gestão de riscos, inferência Bayesiana.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

An insurance policy is a contract that specifies the losses an insurer is legally required to pay in exchange for an initial premium. These losses result from unexpected events, such as car crashes or property damages. They form the basis of claims filed by insureds. This essentially means that the insurance industry sells promises instead of products.

In this context, one of the most concerning problems in this industry is how many of those promises will be materialized in financial losses. In other words, how much money will be needed in order to honor the contracts and keep the company solvent. Additionally, despite the fact that most policies cover a 12-month period, some claims might take several years to be fully paid. Thus, insurance companies rely heavily on historical data to make predictions about their future losses.

According to Charpentier (2014), one example where it is complex to promptly know how much a claim is going to cost involves cases of asbestos liability, especially those arising from prolonged exposures and resulting in lung damages.

That being said, even though historically deterministic methods have been the first ones used to estimate insurance reserves, the increasing availability of computational power has motivated the use of more sophisticated stochastic methods. For example, Verrall (1996), Renshaw and Verrall (1998), Haberman and Renshaw (1996), have considered the use of log-linear models, such as the analysis of variance, whereas De Jong and Zehnwirth (1983), Verrall (1989) and Ntzoufras and Dellaportas (2002) have applied state-space models for loss reserving.

In spite of the fact that there is a vast literature addressing loss reserving, most models developed so far concern the calculation of conditional means as estimates for future claim values. This approach can sometimes fail to provide a good representation of the claims payment distribution, since these measures suffer from a lack of robustness in the presence of outliers.

Therefore, when considering financial modeling there is as much interest in the center of the distribution, as it is in the occurrence of extremes (Frees, 2009). Thus, more robust models for loss reserving, such as in Pitselis et al. (2015), have been established over time. However, those models keep focusing on the computation of conditional means, which still leaves room for the development of models that are able to concomitantly capture the center and the tails of the distribution.

1.1 Dissertation Proposal

For the reasons aforementioned, in this work, an expansion of traditional mean-focused models will be proposed. Extending the work of Choy et al. (2016), the log analysis of variance, the log analysis of covariance and the log state-space are going to receive a Bayesian quantile structure, which will be achieved by allowing residuals to follow an asymmetric Laplace distribution.

This choice is due to the fact that the asymmetric Laplace distribution, with density

$$p(y|\mu,\sigma) = \frac{p(1-p)}{\sigma}\exp\Big\{-\frac{1}{\sigma}\rho_p(y-\mu)\Big\}, \tag{1.1}$$

where $y \in \mathbb{R}, \sigma > 0, p \in (0,1)$, $\mu \in \mathbb{R}$, and $\rho_p(u) = u[p - I(u < 0)]$, with $I(.)$ being the indicator function, has the following property

$$\int_{-\infty}^{\mu} p(y|\mu,\sigma)dy = p. \tag{1.2}$$

Therefore, the exercise of fixing p and estimating μ will provide a conditional quantile of this distribution. This approach allows not only the prediction of central reserves estimates, such as the median, but also the projection of a region where future observations are more likely to be, by forecasting extreme quantiles. The modeling process will adopt the theoretical framework provided by Yu and Moyeed (2001), Tsionas (2003) and Kozumi and Kobayashi (2011). The asymmetric Laplace will be introduced in chapter 3.

According to Bernardi et al. (2016), this pathway is usually adopted when modeling the tail behavior of the underlying distribution is of primary interest. Therefore, once with the advent of Solvency II [1] regulations in the insurance market have become tighter worldwide, knowing the tail behavior of the claims distribution should be of particular concern. Once it could assist companies in attaining legal requirements, while still protecting themselves against unexpected risks, such as extreme losses.

Finally, the Bayesian approach provides a flexible way, with relative computational ease, to explore the entire parametric space, compensating for uncertainty in the estimation process. Also, it conveniently provides predictive distributions, which are the main objective of reserves forecasting.

1.2 Structure

This dissertation is organized as follows. Chapter 2 presents a brief discussion of the fundamentals of Bayesian inference, exposing topics such as point and interval estimation, Markov chain Monte Carlo algorithms and it finishes with a description of model selection criteria.

Chapter 3 brings the principal concepts used this work. It describes run-off triangles and how the pattern evolves through time. Also, the basic model for run-off triangles is described. In addition, this chapter introduces the asymmetric Laplace distribution, ending with concise discussions on fixed-effects models and the dynamic linear model.

[1] see https://eiopa.europa.eu/regulation-supervision/insurance/solvency-ii

In chapter 4 the traditional models for loss reserving are presented and the Bayesian inferential procedure is developed. This chapter ends with an application, where simulations are carried out and the results from Choy et al. (2016) are reproduced.

Finally, chapter 5 introduces the models with asymmetric Laplace distribution (ALD). At the end, the Bayesian inferential procedure for these models is developed and the results are compared with the traditional models used for loss reserving. Lastly, chapter 6 brings the final remarks.

Chapter 2

Fundamentals of Bayesian Inference

2.1 Introduction

This chapter introduces some basic elements of Bayesian inference which are going to be used throughout this dissertation. Sections 2.2 and 2.3 show elementary topics of Bayesian statistics, including the process for obtaining posterior and predictive distributions, and estimation procedures through the Bayesian viewpoint. Section 2.4 describes Markov chain Monte Carlo simulation methods. Finally, section 2.5 addresses the issue of model selection and briefly discusses the Deviance Information Criterion and Watanabe-Akaike Information Criterion. For more details about Bayesian inference, see Migon et al. (2014), Gamerman and Lopes (2006), Gelman et al. (2013), Gelman et al. (2014), Rizzo (2007).

2.2 Bayesian Inference

The Bayesian approach is a sort of inferential procedure, which combines data-driven likelihoods with prior beliefs about the phenomenon under study. Then, the process of blending information is updated through the posterior distribution.

Let $\boldsymbol{\theta}$ be an unknown parameter of interest modeled as a random quantity, and $\boldsymbol{\Theta}$ a parametric space, such that $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Suppose that all the prior belief about $\boldsymbol{\theta}$, previously to data observation, can be quantified by $p\left(\boldsymbol{\theta}\right)$. Also, let $\mathbf{Y}$ be a random variable describing the phenomenon under study, and $\mathbf{y}$ be a particular realization of $\mathbf{Y}$. The likelihood function, denoted by $\mathcal{L}\left(\boldsymbol{\theta};\mathbf{y}\right)$ expresses all the information coming from the data about $\boldsymbol{\theta}$. The likelihood function has the same form as the probability model fitted to $\mathbf{Y}$, denoted by $p\left(\mathbf{y}|\boldsymbol{\theta}\right)$, but it is seen as a function of $\boldsymbol{\theta}$ conditional to the vector of observations $\mathbf{y}$.

Combining prior information, $p\left(\boldsymbol{\theta}\right)$, with the data, $p\left(\mathbf{y}|\boldsymbol{\theta}\right)$, the knowledge about $\boldsymbol{\theta}$ is updated. The posterior distribution, $p\left(\boldsymbol{\theta}|\mathbf{y}\right)$, summarizes the current state of knowledge about $\boldsymbol{\theta}$ given the data.

The posterior distribution, $p\left(\boldsymbol{\theta}|\mathbf{y}\right)$, can be obtained by using the Bayes' Theorem, or Bayes' Rule, which is formulated as follows:

$$p\left(\boldsymbol{\theta}|\mathbf{y}\right) = \frac{p\left(\mathbf{y}|\boldsymbol{\theta}\right)\ p\left(\boldsymbol{\theta}\right)}{p\left(\mathbf{y}\right)}, \tag{2.1}$$

where, $p(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$ has a closed form in only few cases. Thus, as it does not depend on $\boldsymbol{\theta}$, it is common to represent 2.1 as

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\,. \tag{2.2}$$

It is now clear that under the Bayesian scope the parameters of interest also receive a random treatment. This fact makes this type of statistical inference very powerful and appealing, since it is possible to explore the entire parametric space and summarize this result by a probability distribution.

2.2.1 Predictive Distributions

Before the data $\mathbf{y}$ are observed, the marginal, or prior predictive, distribution of the data is

$$p(\mathbf{y}) = \int p(\mathbf{y}, \boldsymbol{\theta})\, d\boldsymbol{\theta} = \int p(\boldsymbol{\theta})\, p(\mathbf{y}|\boldsymbol{\theta})\, d\boldsymbol{\theta}. \tag{2.3}$$

After the data $\mathbf{y}$ have been observed, the same procedure can be used to make inferences about an unknown observable quantity, $\tilde{\mathbf{y}}$, following the same generating process. The distribution of $\tilde{\mathbf{y}}$ is called posterior predictive, because it is conditional on the observed data $\mathbf{y}$:

$$\begin{aligned} p(\tilde{\mathbf{y}}|\mathbf{y}) &= \int p(\tilde{\mathbf{y}}|\mathbf{y}, \boldsymbol{\theta})\, d\boldsymbol{\theta} \\ &= \int p(\tilde{\mathbf{y}}|\boldsymbol{\theta}, \mathbf{y})\, p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta} \\ &= \int p(\tilde{\mathbf{y}}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta}. \end{aligned} \tag{2.4}$$

The last step follows from the assumption that $\mathbf{y}$ and $\tilde{\mathbf{y}}$ are conditionally independent given the parameter $\boldsymbol{\theta}$.

2.3 Estimation

2.3.1 Point Estimation

Estimation is one of the central problems in statistics. A good estimator should have a near-zero error, this means that it has to be as near to the true value of the parameter as possible. In other words, the best estimator $\hat{\boldsymbol{\theta}}$ is a function of data that yields the result closest to $\boldsymbol{\theta}$.

In decision theory, one assumes that for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and every course of action $\boldsymbol{\delta} \in \boldsymbol{\Theta}$ there is a loss function $L(\boldsymbol{\theta}, \boldsymbol{\delta})$ that quantifies the loss of taking decision $\boldsymbol{\delta}$ when $\boldsymbol{\theta}$ occurs. Then, one can proceed to examine, between all actions considered, which one has the lowest risk, or average loss. The risk, $R(\boldsymbol{\delta})$ is defined as

$$R(\boldsymbol{\delta}) = E_{\boldsymbol{\theta}|\boldsymbol{\delta}}\left[L(\boldsymbol{\theta}, \boldsymbol{\delta})\right] = \int_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{\delta})\, p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta}. \tag{2.5}$$

Therefore, a decision $\boldsymbol{\delta}^*$ is optimal if it has minimum risk, namely $R(\boldsymbol{\delta}^*) < R(\boldsymbol{\delta}), \forall \boldsymbol{\delta}$. The is called the Bayes rule, and its risk the Bayes risk. In practice, estimators are obtained by minimizing the expected posterior risk for a particular loss function. Most loss functions are symmetric, the most common ones and their estimators are listed bellow.

1. Absolute Loss Function : $L(\boldsymbol{\theta}, \boldsymbol{\delta}) = |\boldsymbol{\delta} - \boldsymbol{\theta}|$

   Its estimator is the posterior median: $\hat{\boldsymbol{\theta}} = \int_{-\infty}^{\hat{\boldsymbol{\theta}}} p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta} = 0.5;$

2. Quadratic Loss Function: $L(\boldsymbol{\theta}, \boldsymbol{\delta}) = (\boldsymbol{\delta} - \boldsymbol{\theta})^T (\boldsymbol{\delta} - \boldsymbol{\theta})$

   Its estimator is the posterior mean: $\hat{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|\mathbf{y}]\,;$

3. 0-1 Loss Function: $L(\boldsymbol{\theta}, \boldsymbol{\delta}) = \begin{cases} 1, \text{ if } |\boldsymbol{\delta} - \boldsymbol{\theta}| > k, \text{ where } k \to 0 \\ 0, \text{ otherwise} \end{cases}$

   Its estimator is the posterior mode: $\hat{\boldsymbol{\theta}} = sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}}\, p(\boldsymbol{\theta}|\mathbf{y})\,.$

2.3.2 Interval Estimation

In spite of the idea of point estimation through the specification of loss functions being a very appealing one, its use reduces all the information provided by the posterior distribution into a single quantitative measure. Thus, it is important to have a figure to quantify how precise is that estimate.

This goal is usually attained by extracting a range of values from the posterior distribution of the unknown $\boldsymbol{\theta}$ and attaching a probability to it. Generally, it is desirable to have a range of values that is as small as possible, but at the same time contains as much probability as possible.

From this desire, comes the idea of Credible Intervals, which is described as follows:

Definition 1. *Let $\boldsymbol{\theta}$ be a quantity defined in $\boldsymbol{\Theta}$. A set $C \subset \boldsymbol{\Theta}$ is a $100(1-\alpha)\%$ credible interval or Bayesian confidence region for $\boldsymbol{\theta}$ if $P(\boldsymbol{\theta} \in C|\boldsymbol{y}) \geq 1 - \alpha$, where $1 - \alpha$ is the credibility or confidence level.*

Therefore, it becomes clear that the Credible Interval is simply a probability evaluation over the posterior distribution of $\boldsymbol{\theta}$. In general, one would want α and C to be as small as possible, implying the posterior to be highly concentrated. Therefore, in order to obtain the shortest intervals, one would want to include in it the points where the posterior density is the highest. This brings up another concept, the Highest Posterior Density Interval (HPDI).

Definition 2. *A $100(1-\alpha)\%$ highest posterior interval for $\boldsymbol{\theta}$ is the $100(1-\alpha)\%$ credible interval C given by $C = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : p(\boldsymbol{\theta}|\boldsymbol{y}) \geq \boldsymbol{k}(\alpha)\}$, where $\boldsymbol{k}(\alpha)$ is the largest constant such that $P(\boldsymbol{\theta} \in C|\boldsymbol{y}) \geq 1 - \alpha$.*

It is important to point out that, even though Credible Intervals are invariant to $1 - \text{to} - 1$ transformations, HPD intervals are not. Therefore, when a parametric transformation is carried out, the interval loses its highest posterior property.

2.4 Markov chain Monte Carlo Methods

Markov chain Monte Carlo (MCMC) methods are a series of techniques used to draw samples from complex distributions. During the development of Bayesian statistics, the problems start becoming more complicated and intractable posterior distributions more common. This in turn, resulted in a necessity to find methods that were able to draw samples from those distributions. Therefore, borrowing concepts already developed by the physical sciences, MCMC methods began to spread through the world of statistics, specially in Bayesian inference problems.

These methods work by simulating a Markov chain, that will converge to the target distribution by meeting some mild requirements. A Markov chain, $\{\boldsymbol{\theta}^{(t)}\}$, is a sequence of dependent random variables

$$\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(t)}, \ldots$$

such that the conditional probability distribution of $\boldsymbol{\theta}^{(t)}$, given all the past variables, depends only on $\boldsymbol{\theta}^{(t-1)}$. In order for a Markov chain to have a stationary distribution it has to meet the requirements of being irreducible, aperiodic, recurrent. The algorithms discussed next almost always generate Markov chains with those traits.

2.4.1 Metropolis-Hastings Algorithm

The Metropolis-Hastings (MH) algorithm basically works by simulating a candidate value from a proposal distribution, and accepting it, as coming from the target, according to a probability provided by the acceptance function, $\alpha\left(\boldsymbol{\theta}^{(cand)}|\boldsymbol{\theta}^{(t\text{-}1)}\right)$. This function plays the role of balancing two constraints. The first one is that the sampler should be able to visit higher probability areas of the target distribution. The second one is that the sampler should explore the space and avoid getting stuck at one point.

Thus, let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ be a parametric vector, and suppose one wants to generate a sample $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(t)}$ from the limiting (target) distribution of a Markov chain, $p\left(\boldsymbol{\theta}\right)$, which is known up to a normalizing constant, using the proposal $q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)})$. The Metropolis-Hastings algorithm would have the following description:

Algorithm 1

1. At $t = 0$ set $\boldsymbol{\theta}^{(0)}$, as a starting point for the chain;
2. Generate candidate $\boldsymbol{\theta}^{(cand)} \sim q\left(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t\text{-}1)}\right)$;
3. Calculate acceptance probability $\alpha\left(\boldsymbol{\theta}^{(cand)}|\boldsymbol{\theta}^{(t\text{-}1)}\right) = min\left\{1, \frac{q\left(\boldsymbol{\theta}^{(t\text{-}1)}|\boldsymbol{\theta}^{(cand)}\right)p\left(\boldsymbol{\theta}^{(cand)}\right)}{q\left(\boldsymbol{\theta}^{(cand)}|\boldsymbol{\theta}^{(t\text{-}1)}\right)p\left(\boldsymbol{\theta}^{(t\text{-}1)}\right)}\right\}$;
4. Generate $u \sim Uniform\left(0, 1\right)$;
5. If $u < \alpha$, then $\boldsymbol{\theta}^{(t)} \leftarrow \boldsymbol{\theta}^{(cand)}$, otherwise $\boldsymbol{\theta}^{(t)} \leftarrow \boldsymbol{\theta}^{(t\text{-}1)}$;
6. Repeat step 2 to 5 until convergence is reached.

2.4.2 Gibbs Algorithm

The Gibbs sampler is a particular case of the Metropolis-Hastings, where the proposed values are always accepted. This algorithm works by generating posterior samples of each variable, or block of variables, conditional to the current value of all the other components of the parametric vector.

Again, let us consider $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ as being a vector of parameters of dimension d, $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_d\}$. Denoting θ_i as the i^{th} variable, then, the full conditional posterior distribution of θ_i will be indicated by $p\left(\theta_i|\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_d, \mathbf{y}\right)$. Therefore, the Gibbs algorithm will have the following description:

Algorithm 2

1. At $t = 0$ set initial values $\boldsymbol{\theta}^{(0)} = \{\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_d^{(0)}\}$;
2. Obtain new values $\boldsymbol{\theta}^{(t)} = \{\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_d^{(t)}\}$ by random sampling from the full conditional posterior distribution of each variable:

$$\begin{aligned}
&\theta_1^{(t)} \sim p\left(\theta_1|\theta_2^{(t-1)}, \ldots, \theta_d^{(t-1)}, \mathbf{y}\right)\\
&\theta_2^{(t)} \sim p\left(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \ldots \theta_d^{(t-1)}, \mathbf{y}\right)\\
&\vdots\\
&\theta_d^{(t)} \sim p\left(\theta_d|\theta_1^{(t)}, \ldots, \theta_{d-1}^{(t)}, \mathbf{y}\right)
\end{aligned}$$

3. Repeat the previous step until convergence is reached.

2.5 Model Selection

Model selection is the exercise of selecting a statistical model from a set of candidates. This task is carried out by checking the in-sample prediction accuracy of the model. For historical reasons, the measures used to aid the process of model selection are usually called *information criteria*. These measures are typically based on the deviance, which is defined as the data log predictive density conditional to a posterior point estimate, multiplied by -2; that is $D(\hat{\boldsymbol{\theta}}) = -2\ log\ p(\mathbf{y}|\hat{\boldsymbol{\theta}})$.

2.5.1 Deviance Information Criterion

The Deviance Information Criterion (DIC) could be considered a Bayesian version of the well-known Akaike Information Criterion (AIC), which is calculated according to

$$AIC = -2\ log\ p\left(\mathbf{y}|\hat{\boldsymbol{\theta}}_{MLE}\right) + 2k, \tag{2.6}$$

where the maximum likelihood estimate, $\hat{\boldsymbol{\theta}}_{MLE}$, would be replaced by the posterior mean $\hat{\boldsymbol{\theta}}_{Bayes} = E\left(\boldsymbol{\theta}|\mathbf{y}\right)$, and k, the number of parameters in the model, would be replaced by a measure representing the effective number of parameters in the model, p_{DIC}. Therefore, the DIC is given by

$$DIC = -2\ log\ p\left(\mathbf{y}|\hat{\boldsymbol{\theta}}_{Bayes}\right) + 2p_{DIC}, \tag{2.7}$$

where the quantity p_{DIC} is calculated in terms of an average of $\boldsymbol{\theta}$ over its posterior distribution. Then, having a posterior distribution sample of size T, and letting $\boldsymbol{\theta}^{(t)}$ be the t^{th} iteration, for $t = 1, 2, \ldots, T$,

$$P_{DIC} = 2\left(log\ p\left(\mathbf{y}|\hat{\boldsymbol{\theta}}_{Bayes}\right) - \frac{1}{T}\sum_{t=1}^{T} log\ p\left(\mathbf{y}|\boldsymbol{\theta}^{(t)}\right)\right). \tag{2.8}$$

2.5.2 Watanabe-Akaike Information Criterion

The Watanabe-Akaike Information Criterion (WAIC) could be seen as a fully Bayesian alternative to other information criteria, because it does not start by ignoring the posterior distribution uncertainty by using a point estimate, like DIC does. Also, both the log posterior predictive density and the correction for effective number of parameters are pointwise.

Suppose there is a sample of n data points, $\mathbf{y} = \{y_1, y_2, \ldots, y_n\}$, and T MCMC posterior samples, $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(T)}$. Firstly, it proceeds by calculating the log pointwise posterior predictive density *lppd* as the following

$$lppd = \sum_{i=1}^{n} log\left(\frac{1}{T}\sum_{t=1}^{T} p\left(y_i|\boldsymbol{\theta}^{(t)}\right)\right), \tag{2.9}$$

then a measure to adjust for overfitting, p_{WAIC}, is obtained by

$$p_{WAIC_1} = 2\sum_{i=1}^{n}\left(log\left(\frac{1}{T}\sum_{t=1}^{T} p\left(y_i|\boldsymbol{\theta}^{(t)}\right)\right) - \frac{1}{T}\sum_{t=1}^{T} log\ p\left(y_i|\boldsymbol{\theta}^{(t)}\right)\right) \tag{2.10}$$

or

$$p_{WAIC_2} = \sum_{i=1}^{n} V_{t=1}^{T}\left(log\ p\left(y_i|\boldsymbol{\theta}^{(t)}\right)\right). \tag{2.11}$$

Where $V_{t=1}^{T}$ represents the individual sample variance for each data point, for instance, for a random variable a, $V_{i=1}^{n} a_i = \frac{1}{n-1}\sum_{i=1}^{n}(a_i - \bar{a})$. Therefore, the WAIC is defined as

$$WAIC = -2\left(lppd - p_{WAIC}\right). \tag{2.12}$$

Chapter 3

Introductory Concepts

3.1 Introduction

This chapter brings the most important concepts used in this work. Section 3.2 describes the concept of a run-off triangle and a derivation of a basic model for this particular data structure is demonstrated. Section 3.3 presents the student's t representation as scale mixture of normal distributions. Section 3.4 introduce the asymmetric Laplace distribution along with a mixture representation that enables the implementation of Gibbs sampling in a quantile regression structure. Finally sections 3.5 and 3.6 detail the two classes of models used in this work.

3.2 Run-off Triangles

Loss reserving data are usually presented as a triangle structure called run-off. In the case of losses Incurred but Not Reported (IBNR), it can display the policy-year i in lines and the lag-year j in columns. This run-off layout is showed in Table 3.1, where $Y_{ij},\ i = 1, \ldots, n;\ j = 1, \ldots, n-i+1$, is the amount paid by the insurance company for claims occurred in year i (policy year i) and settled in year $j-1$ (lag year j). Note that each Y_{ij} represents incremental claims, instead of cumulative claims. Thus, statistical models have been developed to forecast the lower triangle and these values will constitute the reserves of the insurance company.

For instance, if claims until year n have been observed, the reserves for the next year will be the predicted values right below the last diagonal of observed claims. For example Table 3.2

Table 3.1: Run-off triangle for claim amount data.

		Lag Year j				
		1	2	$\ldots$	$n-1$	n
	1	$Y_{1,1}$	$Y_{1,2}$	$\ldots$	$Y_{1,n-1}$	$Y_{1,n}$
Policy	2	$Y_{2,1}$	$Y_{2,2}$	$\ldots$	$Y_{2,n-1}$	?
Year i	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$n-1$	$Y_{n-1,1}$	$Y_{n-1,2}$	?	?	?
	n	$Yn,1$	?	?	?	?

Table 3.2: Diagonal of run-off triangle for year $n+1$.

		Lag Year j				
		1	2	...	$n-1$	n
	1	$Y_{1,1}$	$Y_{1,2}$	...	$Y_{1,n-1}$	$Y_{1,n}$
Policy	2	$Y_{2,1}$	$Y_{2,2}$	...	$Y_{2,n-1}$	$Y_{2,n-1}$
Year i	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$n-1$	$Y_{n-1,1}$	$Y_{n-1,2}$	$Y_{n-1,3}$	?	?
	n	$Yn,1$	$Y_{n,2}$	?	?	?
	$n+1$	$Y_{n+1,1}$	?	?	?	?

shows marked in red the estimated claims of year $n+1$. It is important to notice that these triangles can be built on different time frames, for instance, it could be monthly rather than annual. Refer to Schmidt and Seminar (2006) for more information on run-off triangles.

3.2.1 Basic Model for Run-Off Triangles

This subsection is a brief discussion of the basic model for run-off triangles (Verrall, 1989; Kremer, 1982) and its simplest formulation. Models in subsequent chapters will follow the same essential foundations.

Let $\{Y_{ij}|j \leq n-i+1,\ i \leq n\}$ be the history of claim payments available to the company after business has been running for n years. The triangles take the form shown in table 3.1 and it is assumed that $Y_{ij} > 0$, $\forall\ i, j$.

Then, let U_i and S_j be mutually independent parameters representing row i and column j, respectively. Also, let R_{ij} be the random errors with expectation equals to one. The basic model for a run-off triangle would have the following multiplicative form

$$Y_{ij} = U_i S_j R_{ij}. \tag{3.1}$$

A natural step now is to take the logarithm of equation 3.1 in order to have a linear representation as follows

$$Z_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad \forall j \leq n-i+1,\ \forall i \leq n. \tag{3.2}$$

Where,

$$\begin{aligned}
Z_{ij} &= \log\left(Y_{ij}\right),\ \epsilon_{ij} = \log\left(R_{ij}\right), \\
\alpha_i &= \log\left(U_i\right) - \frac{1}{n}\sum_{l=1}^{n} \log\left(U_l\right), \\
\beta_j &= \log\left(S_j\right) - \frac{1}{n}\sum_{l=1}^{n} \log\left(S_l\right), \\
\mu &= \frac{1}{n}\sum_{l=1}^{n} \left(\log(U_l) + \log(S_l)\right).
\end{aligned} \tag{3.3}$$

also,
$$\sum_{i=1}^{n} \alpha_i = \sum_{j=i}^{n} \beta_j = 0. \tag{3.4}$$

Equation 3.4 tackles the problem of model identifiability, which is a common issue in every model used in this work. A model is identifiable if the parameter values are able to uniquely determine the distribution of the data. This problem usually happens when the model is poorly specified, therefore those constraints presented above will be essential during the implementation phase of this work. The final assumption is that the errors, ϵ_{ij}, $\forall j \leq n - i + 1$, $\forall i \leq n$, are uncorrelated with expectation equals to zero and constant variance equals to σ^2.

All models used in this work are going to follow some variation of equation 3.2. They are going to be described in the next two chapters.

3.3 The Student's t distribution as a scale mixture of normal distributions

In insurance applications it is well-known that data sets, in most cases, contain irregular and extremes values which makes the assumption of normality not very realistic. However, in the Bayesian framework, working with more complex distributions can make implementations very difficult. It has been shown by Choy et al. (2016) that using scale mixture of the normal distribution can alleviate that burden.

Therefore, if a continuous random variable X has a student's t distribution with location $-\infty < \mu < \infty$, scale $\sigma^2 > 0$ and degrees of freedom $\nu > 0$, its density is represented by

$$t_\nu \left(X|\mu, \sigma^2\right) = \int_0^\infty \mathcal{N}\left(X|\mu, \lambda^{-1}\sigma^2\right) \Gamma\left(\lambda|\frac{\nu}{2}, \frac{\nu}{2}\right) d\lambda, -\infty < x < \infty. \tag{3.5}$$

where $\mathcal{N}(.|.,.)$ and $\Gamma(.|.,.)$ are the normal and gamma distributions, respectively. Using this representation X has, conditional on λ, $\mathcal{N}\left(X|\mu, \lambda^{-1}\sigma^2\right)$ distribution, with λ being the scale mixture parameter. It can be seen in Choy et al. (2016) that λ can also be useful for outliers detection. For more information on normal scale mixtures see Andrews and Mallows (1974), Choy and Chan (2003) and Qin et al. (2003).

3.4 The Asymmetric Laplace Distribution

This section presents the asymmetric Laplace distribution and its mixture representation. This distribution allows the performance of quantile analysis, being particularly useful in the Bayesian inferential process.

Definition 3. *A random variable Y is said to have an asymmetric Laplace distribution, denote by $Y \sim ALD(\mu, \sigma, p)$, if its density function is given by*

$$p(y|\mu, \sigma, p) = \frac{p\,(1-p)}{\sigma} \exp\left\{ -\frac{(y-\mu)}{\sigma}\left[p - I(y \leq 0)\right] \right\} I\left(-\infty < y < \infty\right), \tag{3.6}$$

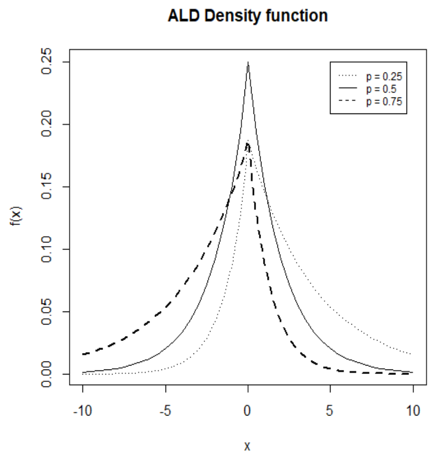


Figure 3.1: Asymmetric Laplace Distribution with $\mu = 0$ and $\sigma = 1$.

where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter, $0 < p < 1$ is the skew parameter, and $I(.)$ is the indicator function. This distribution is skewed to the left if $p > 0.5$, skewed to the right if $p < 0.5$ and symmetric if $p = 0.5$, in this last case the distribution is named Laplace double exponential. Figure 3.1 shows an example of the probability density function of this distribution.

The mean and variance of Y are given by

$$E(Y) = \mu + \frac{\sigma\,(1-2p)}{p(1-p)}, \quad Var(Y) = \frac{\sigma^2(1-2p+2p^2)}{(1-p)^2p^2}. \tag{3.7}$$

For more information on the asymmetric Laplace distribution refer to Yu and Zhang (2005).

Next, the mixture representation of the ALD distribution is described. It will be useful in the Bayesian inferential process discussed in chapter 5 because it allows the implementation of Gibbs sampling. The representation shown here is taken from Kozumi and Kobayashi (2011).

Consider the linear model

$$y_t = \mu + \epsilon_t, \quad t = 1, \ldots, n, \tag{3.8}$$

following 3.6, if we consider $\epsilon_t \sim ALD\,(0, 1, p)$, its density will be given by

$$f(\epsilon_t) = p(1-p)\exp\{-\epsilon_t(p - I_{\epsilon_t<0})\}, \tag{3.9}$$

also, the mean and variance are defined by the expressions in equation 3.7. Therefore, a mixture representation based on standard normal and standard exponential distributions is adopted.

Proposition 1. *Let Z be an exponential variable, with scale σ, and U a standard normal variable. If a random variable ϵ follows an Asymmetric Laplace distribution with density* 3.9*, then we can represent ϵ as a location-scale mixture of normals given by*

$$\epsilon = \theta z + \tau\sqrt{z}u, \tag{3.10}$$

where,

$$\theta = \frac{1-2p}{p(1-p)} \quad \textit{and} \quad \tau^2 = \frac{2}{p(1-p)}. \tag{3.11}$$

Then, equation 3.8 can be written as

$$y_t = \mu + \theta z_t + \tau\sqrt{z_t}u_t. \tag{3.12}$$

In the applications done in this work, in order to allow inference with a scale parameter $\sigma \neq 1$, equation 3.12 will be rewritten as

$$y_t = \mu + \sigma\theta z_t + \sigma\tau\sqrt{z_t}u_t. \tag{3.13}$$

Finally, to allow Gibbs sampling implementation for all parameters, a slight reparametrization is necessary, resulting in 3.14, where $\nu_t = \sigma z_t$,

$$y_t = \mu + \theta\nu_t + \tau\sqrt{\sigma\nu_t}u_t. \tag{3.14}$$

A proof of proposition 1 is available in appendix 3.A.

3.5 Fixed-Effects Models

The fixed-effects model (Montgomery, 2017) is widely adopted in a loss reserving context. These models can be seen as stochastic generalizations of a deterministic method called Chain-Ladder, see Renshaw and Verrall (1998). The basic two-way fixed-effect model without an interaction factor has the representation:

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \begin{cases} i = 1, \dots, a \\ j = 1, \dots, b. \end{cases} \tag{3.15}$$

Where a is usually the number of treatments and b is the number of blocks. In the loss reserving context, i and j will represent, respectively, lines and columns of a run-off triangle. Also, $a = n$, where n is the number of years with observations available, and $b = a - i + 1$. The model in equation 3.15 has an intuitive appeal, where μ is a constant and the effects α_i and β_j represent deviations from μ. This model could also be represented with a mean function $\mu_{ij} = \mu + \alpha_i + \beta_j$, this way 3.15 would simplify to

$$y_{ij} = \mu_{ij} + \epsilon_{ij} \begin{cases} i = 1, \dots, a \\ j = 1, \dots, b. \end{cases} \tag{3.16}$$

The assumptions of this model are basically the presence of constant variance, errors being independent and identically distributed. Also, the α_i, for $i = 1, \dots, a$ and β_j, for $j = 1, \dots, b$ parameters, for indentifiability reasons, are subject to the constraint $\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = 0$. The latter one is a property that enables precise inference to be carried out.

3.6 The State-Space Model

The state-space model is a very general class of models used in a variety of fields, ranging from economics to the natural sciences. These models are basically defined by a pair of equations. Let $\boldsymbol{Y}_t$ be a $(p \times 1)$ vector of observations over time $t = 1, 2, \ldots,$ then:

$$\boldsymbol{Y}_t = \boldsymbol{F}_t' \boldsymbol{\theta}_t + \boldsymbol{v}_t, \tag{3.17}$$

and

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t. \tag{3.18}$$

Where 3.17 is called observation equation and 3.18 is the evolution equation. Also,

- $\boldsymbol{F}_t$ is a known $(n \times p)$ matrix called design matrix;
- $\boldsymbol{\theta}_t$ is a $(n \times 1)$ vector of states at time t;
- $\boldsymbol{G}_t$ is a known $(n \times n)$ matrix called evolution matrix;

Moreover, $\boldsymbol{v}_t$ is a $(p \times 1)$ vector of random noise, mutually independent, with mean zero and covariance matrix $\boldsymbol{V}_t$, $\boldsymbol{w}_t$ is a random noise vector as well, with dimension $(n \times 1)$, mutually independent with mean zero and covariance matrix $\boldsymbol{W}_t$. Besides, the following assumptions are usually made:

- The initial state $\boldsymbol{\theta}_0$ has mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$;
- The errors $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ are uncorrelated among themselves and the initial state, meaning:

$$E(\boldsymbol{v}_t \boldsymbol{w}_t') = \boldsymbol{0}, \text{ and}$$

$$E(\boldsymbol{v}_t \boldsymbol{\theta}_0') = \boldsymbol{0}, \; E(\boldsymbol{w}_t \boldsymbol{\theta}_0') = \boldsymbol{0}, \forall t.$$

Finally, $\boldsymbol{F}_t$ and $\boldsymbol{G}_t$ are non-stochastic matrices, this means if there is a change in time, that change will be specified beforehand. For more on linear state-space models see West and Harrison (2006).

Appendix

3.A Proof of proposition 1

Proof. Kotz et al. (2012).
Let $\epsilon \sim ALD(0,1,p)$, then its probability density function is given by

$$f(\epsilon) = p(1-p)\exp\{-\epsilon(p - I_{\epsilon<0})\}.$$

On one hand, the characteristic function of ϵ can be obtained as follows

$$\begin{aligned}
\varphi_\epsilon(t) = E(e^{it\epsilon}) &= \int_{-\infty}^{\infty} e^{it\epsilon} f(\epsilon) d\epsilon \\
&= p(1-p)\left[\int_{-\infty}^{0} \exp\{it\epsilon + (1-p)\epsilon\} d\epsilon + \int_{0}^{\infty} \exp\{it\epsilon - p\epsilon\} d\epsilon\right] \\
&= p(1-p)\left\{\left[\frac{\exp\{(it+(1-p))\epsilon\}}{it+(1-p)}\right]_{-\infty}^{0} + \left[\frac{\exp\{(it-p)\epsilon\}}{it-p}\right]_{0}^{\infty}\right\} \\
&= p(1-p)\left\{\left[\frac{1}{it-(1-p)} - 0\right] + \left[0 - \frac{1}{it-p}\right]\right\} \\
&= p(1-p)\left[\frac{1}{(it+(1-p))(p-it)}\right] \\
&= \left[\frac{p^2 - i(1-2p)t + p(1-p)}{p(1-p)}\right]^{-1} \\
&= \left[\frac{1}{p(1-p)}t^2 - i\frac{1-2p}{p(1-p)}t + 1\right]^{-1}.
\end{aligned}$$

On the other hand, let $N \sim \mathcal{E}(1)$ and $U \sim \mathcal{N}(0,1)$. Then, the characteristic function of $\epsilon^* = k_1 N + k_2\sqrt{N}U$ can be expressed as

$$\begin{aligned}
\varphi_{\epsilon^*}(t) = E(e^{it\epsilon^*}) &= E(e^{it(k_1N + k_2\sqrt{N}U)}) \\
&= E\Big[E(\exp\{it(k_1N + k_2\sqrt{N}U)\})|N = \nu\Big] \\
&= \int_0^\infty \exp\{itk_1\nu\} E(\exp\{itk_2\sqrt{\nu}U\}) e^{-\nu} d\nu.
\end{aligned} \tag{3.19}$$

From the result above, we have that $E(\exp\{itk_2\sqrt{\nu}U\}) = \varphi_U(tk_2\sqrt{\nu})$. Besides, if $U \sim \mathcal{N}(0,1)$, then its characteristic function is given by

$$\varphi_U(t) = \exp\Big\{ -\frac{1}{2}t^2 \Big\}.$$

Therefore, by replacing this function in equation 3.19 we have

$$\begin{aligned}
\varphi_{\epsilon^*}(t) &= \int_0^\infty \exp\{itk_1\nu\}\exp\Big\{-\frac{1}{2}t^2k_2^2\nu\Big\}e^{-\nu}d\nu \\
&= \int_0^\infty \exp\Bigg\{-\nu\Bigg(1+\frac{1}{2}t^2k_2^2 - ik_1t\Bigg)\Bigg\}d\nu \\
&= \Bigg[\frac{\exp\Big\{-\nu\Big(1+\frac{1}{2}t^2k_2^2-ik_1t\Big)\Big\}}{-\Big(1+\frac{1}{2}t^2k_2^2-ik_1t\Big)}\Bigg]_0^\infty \\
&= \Bigg[\frac{k_2^2}{2}t^2 - ik_1t + 1\Bigg]^{-1}.
\end{aligned}$$

Thus, in order for the above expression to be equal to the characteristic function of an asymmetric Laplace distribution with location $\mu = 0$, scale $\sigma = 1$ and skew p,

$$k_1 = \frac{1-2p}{p(1-p)} \quad \text{and} \quad k_2^2 = \frac{2}{p(1-p)}.$$

$\square$

Chapter 4

Models with Normal Scale Mixture

4.1 Introduction

This chapter intends to make an exposure of the basic formulations for each model used in this work. The focus will be on the conditional mean estimation in the loss reserving context. Section 4.2 contains a brief exhibition of log-anova and log-ancova basic models in the particular context of forecasting reserves. At the end of section 4.2 the formulation of the log-state-space is introduced. Section 4.3 briefly describes the inferential process in the Bayesian context. Section 4.4 brings the results from simulations and real-world data applications using this models. This chapter ends with a discussion in section 4.5.

4.2 Models with Normal Scale Mixture

4.2.1 Log-Anova

The simplest way to model reserves is by using a two-way analysis of variance (Verrall, 1991), which is a linear model that belongs to the fixed-effects class of models. Let Y_{ij} be the claim amount paid to insureds from the policy-year i with $j-1$ years of delay and $Z_{ij} = log(Y_{ij})$, then, for all $j \leq n-i+1,\ i \leq n$:

$$\begin{aligned} Z_{ij} &= \mu_{ij} + \lambda_{ij}^{-\frac{1}{2}}\epsilon_{ij}, \\ \mu_{ij} &= \mu + \alpha_i + \beta_j, \\ \lambda_{ij} &\sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \\ \epsilon_{ij} &\sim \mathcal{N}\left(0, \sigma^2\right), \end{aligned} \tag{4.1}$$

where $\mathcal{N}(.,.)$ and $\mathcal{G}(.,.)$ are normal and gamma distributions, respectively. According to this representation $Z_{ij}|\mu_{ij}, \sigma^2, \lambda_{ij} \sim \mathcal{N}\left(\mu_{ij}, \lambda_{ij}^{-1}\sigma^2\right)$, where μ_{ij} is called the mean function. Also, this model has the same assumptions previously discussed for the fixed-effects models, in section 3.5, with constraints $\sum_i \alpha_i = \sum_j \beta_j = 0$.

4.2.2 Log-Ancova

The analysis of covariance presumes the existence of a variable which is linearly related to the response. In the loss reserving context, one of the ways this model could be used is by assuming there is a linear effect in the policy-year i, therefore the co-variable in this case would be the index i of the lines in the run-off triangle. This model is very similar to the previous one with a slight change in the mean function,

$$\mu_{ij} = \mu + i\alpha + \beta_j. \tag{4.2}$$

The basic model is going to have the same formulation as 4.1, with the mean function given by 4.2. The assumptions are the same as the ones stated in section 3.5, with the constraint $\sum_j \beta_j = 0$. Indeed, there is the possibility of also having a linear effect in the lag-year j.

4.2.3 Log-State-Space

This model belongs to the dynamic linear model class, described in section 3.6. The most important difference from the previous two models is that this one allows dynamic evolution of the parameters and considers the possibility of having an interaction between policy-year and lag-year, making it easier to identify changes in the run-off pattern. A full mathematical derivation of this model is found in Verrall (1994). For all $j \leq n - i + 1,\ i \leq n$, it has mean function given by,

$$\mu_{ij} = \mu + \alpha_i + \beta_{ij} \tag{4.3}$$

Now, β_j has been replaced by β_{ij}, meaning that the lag-year factors vary according to the policy year. Also, a considerable difference from the previous models is the recursive patterns

$$\begin{aligned} \alpha_i &= \alpha_{i-1} + \epsilon_\alpha, \\ \beta_{ij} &= \beta_{i-1j} + \epsilon_\beta. \end{aligned} \tag{4.4}$$

The equations in 4.4 reflect the belief that α and β evolve through time according to stochastic mechanisms represented by ϵ_α and ϵ_β, where $\epsilon_\alpha \sim \mathcal{N}(0, \sigma^2_\alpha)$, and $\epsilon_\beta \sim \mathcal{N}(0, \sigma^2_\beta)$. Also, for identifiability reasons, $\alpha_1 = 0$ and $\boldsymbol{\beta_{i1}} = \mathbf{0}$, for $i = 1, \ldots, n$. The basic model will have the same formulation as the one expressed by equation 4.1 with mean function given by 4.3. Model assumptions will be equal to the ones describe in section 3.6.

4.3 Bayesian Inference

The first step in the Bayesian inference procedure is to write the likelihood function. Therefore, for all models using normal scale mixture representation we know that the distribution of the observations is $Z_{ij}|\mu_{ij}, \sigma^2, \lambda_{ij} \sim \mathcal{N}\left(\mu_{ij}, \lambda_{ij}^{-1}\sigma^2\right)$. However, due to the presence of $\lambda_{ij},\ \forall i, j$ our representation will be based on the extended likelihood function. Then, let $\boldsymbol{\theta}$ be the vector containing all model parameters, $\boldsymbol{z}$ be the vector with all observations, and $\boldsymbol{\lambda}$ be a vector with all the λ_{ij}, the extended likelihood function will be given by

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{z}) = \prod_{i=1}^{n} \prod_{j=1}^{n-i+1} p(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{\lambda}) = \prod_{i=1}^{n} \prod_{j=1}^{n-i+1} \frac{1}{\sqrt{2\pi \frac{\sigma^2}{\lambda_{ij}}}} \exp\left\{ -\frac{\lambda_{ij}}{2\sigma^2}(z_{ij} - \mu_{ij})^2 \right\} I(\lambda_{ij} > 0) \quad (4.5)$$

where μ_{ij} will be the particular mean function for each model described in section 4.2. That being said, it is now necessary to set prior distributions for each parameter in the model, we will be using here the same priors as the ones in Choy et al. (2016).

4.3.1 Prior Distributions

Set the priors for the log-anova in the normal distribution case as: $\mu \sim \mathcal{N}\left(0, \sigma_\mu^2\right)$; $\sigma^2 \sim \mathcal{IG}\left(a, b\right)$; $\alpha_i \sim \mathcal{N}\left(0, \sigma_\alpha^2\right)$, for $i = 1, \ldots, n-1$; $\beta_j \sim \mathcal{N}\left(0, \sigma_\beta^2\right)$, for $j = 1, \ldots, n-1$. For the log-ancova, once we have only one α, the only difference will be: $\alpha \sim \mathcal{N}\left(0, \sigma_\alpha^2\right)$. And for the log-state-space, due to the recursive patterns, priors for α_i and β_{ij} will be: $\alpha_i \sim \mathcal{N}\left(\alpha_{i-1}, \sigma_\alpha^2\right)$, for $i = 2, \ldots, n$; $\beta_{ij} \sim \mathcal{N}\left(\beta_{i-1j}, \sigma_\beta^2\right)$, for $i = 1, \ldots, n-j+1$, $j = 2, \ldots, n$. Also, once we need to sample a new β_{ij} to make the predictions, for the log-state-space there will be a need to estimate σ_β^2 to which we will assume the following prior: $\sigma_\beta^2 \sim \mathcal{U}\left(0, A\right)$. Where, $\mathcal{N}(.,.)$, $\mathcal{IG}(.,.)$ and $\mathcal{U}(.,.)$ are normal, inverse-gamma and uniform distributions, respectively.

In the case of models with errors following a student's t distribution, the priors assumed for the common parameters were the same as the ones for the normal models, the difference here is the two additional parameters $\boldsymbol{\lambda}$ and ν. We have set the priors $\lambda_{ij} \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ and, following Fonseca et al. (2008), $p\left(\nu\right) \propto \left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}} \left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}$. Here, $\mathcal{G}(.,.)$ represents a gamma distribution, and $\psi'(.)$ is the trigamma function.

Also, it is being considered for the log-anova $\alpha_n = \sum_{i=1}^{n-1} \alpha_i$ and $\beta_n = \sum_{j=1}^{n-1} \beta_j$, and for the log-ancova only the latter applies. The restrictions for the log-state-space have already been state in the previous section.

4.3.2 The Joint Posterior Distribution

The joint posterior distribution of the parameters will be obtained via the methods discussed in chapter 2, particularly by equation 2.2, the Bayes' theorem. The joint posterior for the log-anova is given by

$$p\left(\mu, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \nu|\boldsymbol{z}\right) \propto \left\{ \prod_{i=1}^{n} \prod_{j=1}^{n-i+1} \left(\frac{\lambda_{ij}}{\sigma^2}\right)^{1/2} \exp\left\{ -\frac{\lambda_{ij}}{2\sigma^2}(z_{ij} - \mu - \alpha_i - \beta_j)^2 \right\} I(\lambda_{ij} > 0) \right\} \times$$
$$\left\{ \prod_{i=1}^{n-1} \exp\left\{ -\frac{\alpha_i^2}{2\sigma_\alpha^2} \right\} \right\} \times \left\{ \prod_{j=1}^{n-1} \exp\left\{ -\frac{\beta_j^2}{2\sigma_\beta^2} \right\} \right\} \times \exp\left\{ -\frac{\mu^2}{2\sigma_\mu^2} \right\} \times \left\{ \left(\frac{1}{\sigma^2}\right)^{a+1} \exp\left\{ -\frac{b}{\sigma^2} \right\} \right\} \times$$
$$\left\{ \prod_{i=1}^{n} \prod_{j=1}^{n-i+1} \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \lambda_{ij}^{\frac{\nu}{2}-1} \exp\left\{ -\frac{\lambda_{ij}\nu}{2} \right\} I(\lambda_{ij} > 0) \right\} \times \left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}} \left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right]^{\frac{1}{2}}.$$

where, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{n-1})$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{n-1})$ and $\lambda = (\lambda_{11}, \dots, \lambda_{nn})$.

For the log-ancova there is only a slight modification in the mean function, that is now given by equation 4.2. Also, since α is now a single parameter, its prior distribution will be replaced in the above expression by

$$\exp\left\{-\frac{\alpha^2}{2\sigma_\alpha^2}\right\}.$$

Finally, the joint posterior distribution for the log-state-space will have several alterations, when compared to the previous models. This distribution will be given by the following

$$p\left(\mu, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \nu, \sigma_\beta^2 | \boldsymbol{z}\right) \propto \left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left(\frac{\lambda_{ij}}{\sigma^2}\right)^{1/2}\exp\left\{-\frac{\lambda_{ij}}{2\sigma^2}(z_{ij}-\mu-\alpha_i-\beta_{ij})^2\right\}I(\lambda_{ij}>0)\right\}\times$$
$$\left\{\prod_{i=2}^{n}\exp\left\{-\frac{(\alpha_i-\alpha_{i-1})^2}{2\sigma_\alpha^2}\right\}\right\}\times\left\{\prod_{j=1}^{n}\prod_{i=1}^{n-j+1}\exp\left\{-\frac{(\beta_{ij}-\beta_{i-1j})^2}{2\sigma_\beta^2}\right\}\right\}\times\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\}\times$$
$$\left\{\left(\frac{1}{\sigma^2}\right)^{a+1}\exp\left\{-\frac{b}{\sigma^2}\right\}\right\}\times$$
$$\left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}\lambda_{ij}^{\frac{\nu}{2}-1}\exp\left\{-\frac{\lambda_{ij}\nu}{2}\right\}I(\lambda_{ij}>0)\right\}\times\left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}}\left[\psi'\left(\frac{\nu}{2}\right)-\psi'\left(\frac{\nu+1}{2}\right)-\frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}.$$

where, $\boldsymbol{\alpha} = (\alpha_2, \dots, \alpha_{n-1})$, $\boldsymbol{\beta} = (\beta_{21}, \dots, \beta_{n1})$ and $\lambda = (\lambda_{11}, \dots, \lambda_{nn})$.

Finally, as the above expressions do not have closed forms, Markov chain Monte Carlo methods, as discussed in chapter 2, are used for parameter estimation. In the normal distribution case, an implementation of the Gibbs sampler will be executed as follows:

Gibbs sampling

1. At $t = 0$ set initial values $\boldsymbol{\theta}^{(0)} = \{\mu^{(0)}\sigma^{2(0)}, \boldsymbol{\alpha}^{(0)}\boldsymbol{\beta}^{(0)}\}$;
2. Obtain new values $\boldsymbol{\theta}^{(t)} = \{\mu^{(t)}\sigma^{2(t)}, \boldsymbol{\alpha}^{(t)}\boldsymbol{\beta}^{(t)}\}$ by randomly sampling from:

$$\mu^{(t)} \sim p\left(\mu|\sigma^{2(t-1)}, \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \boldsymbol{z}\right)$$
$$\sigma^{2(t)} \sim p\left(\sigma^2|\mu^{(t)}, \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \boldsymbol{z}\right)$$
$$\boldsymbol{\alpha}^{(t)} \sim p\left(\boldsymbol{\alpha}|\mu^{(t)}, \sigma^{2(t)}, \boldsymbol{\beta}^{(t-1)}, \boldsymbol{z}\right)$$
$$\boldsymbol{\beta}^{(t)} \sim p\left(\boldsymbol{\beta}|\mu^{(t)}, \sigma^{2(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{z}\right)$$

3. Repeat the previous step until convergence is reached,

this representation will receive an additional step for σ_β^2 during the estimation procedure for the log-state-space. Furthermore, α will be a single scalar parameter for the log-ancova.

In the student's t case, there will be the inclusion of $\boldsymbol{\lambda} \sim p(\boldsymbol{\lambda}|\mu, \nu, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z})$, along with a Metropolis Hastings step for ν. In the latter case, the option to use a Metropolis random walk with log-adaptive proposal was made. It follows that:

Metropolis Hastings Random-Walk with log-adaptive proposal

1. At $t = 0$ set $\nu^{(0)}$, $\bar{\nu}$; ${\sigma_m^2}^{(0)} = 2.4^2$, and ${\sigma_0}^{(0)} = 1$.
2. Take a Metropolis Hastings random-walk step with ${\sigma_m^2}^{(0)}$ and ${\sigma_0}^{(0)}$;
3. Calculate acceptance probability $\hat{r}^{(t)}$;
4. Generate $u \sim Uniform\,(0, 1)$;
5. If $u < \hat{r}^{(t)}$, then $\nu^{(t)} \leftarrow \nu^{(cand)}$, otherwise $\nu^{(t)} \leftarrow \nu^{(t\text{-}1)}$;
6. Calculate:

$$\hat{\sigma}_m^{(t)} = \hat{\sigma}_m^{(t-1)}\left(\nu^{(t-1)} - \bar{\nu}^{(t-1)}\right)^2;$$
$$\gamma^{(t)} = \tfrac{1}{t^c};$$
$$\log\left({\sigma_m^2}^{(t)}\right) = \log\left({\sigma_m^2}^{(t-1)}\right) + \gamma^{(t)}\left(\hat{r}^{(t)} - r^{opt}\right);$$
$$\bar{\nu}^{(t)} = \bar{\nu}^{(t-1)} + \gamma^{(t)}\left(\nu^{(t-1)} - \bar{\nu}^{(t-1)}\right);$$
$${\sigma_0}^{(t)} = {\sigma_0}^{(t-1)} + \gamma^{(t)}(\left(\nu^{(t-1)} - \bar{\nu}^{(t-1)}\right)^2 - {\sigma_0}^{(t-1)})$$

7. Repeat step 2 to 6 until convergence is reached.

The proposal is $Q(\nu|.) \sim \mathcal{N}(\nu, \sigma_m\sigma_0)$, with σ_m and σ_0 being parameters of this distribution. Also, $\hat{r}$ stands for the acceptance rate, and r^{opt} for the optimal theoretical acceptance rate of the MH, which the authors state to be 0.234. And $\gamma^{(t)}$ is a deterministic sequence satisfying $\sum_{t=1}^{\infty}\gamma^{(t)} = \infty$ and $\sum_{t=1}^{\infty}\gamma^{2(t)} < \infty$, c was set to 0.8 (Shaby and Wells, 2010).

Finally, for the log-state-space model with normal and Student t distributions, the variance of β_{ij}, $\forall i, j$ was included in the Gibbs sampler step. The derivation of full conditional posterior distributions used above is available in appendix 4.A.

4.3.3 Predictions

The target of this process is the predictive distribution $p\left(z_{ij}^*|z_{ij}\right)$, where z_{ij}^* is the unobserved data in the lower triangle and z_{ij} represent the observations available (Vilela, 2013). Therefore, let $\boldsymbol{\theta}$ be the vector with all parameters, except λ_{ij} and let λ_{ij}^* be the scale-mixture parameters associated with the unobserved z_{ij}^*, $\forall i, j$. Then, the predictive would be

$$\begin{aligned}
p(z_{ij}^*|z_{ij}) &= \int p(z_{ij}^*, \lambda_{ij}^*, \boldsymbol{\theta}, \lambda_{ij}|z_{ij})d\boldsymbol{\theta}d\lambda_{ij}d\lambda_{ij}^* \\
&= \int p(z_{ij}^*|\lambda_{ij}^*, \boldsymbol{\theta}, \lambda_{ij}, z_{ij})p(\lambda_{ij}^*, \boldsymbol{\theta}, \lambda_{ij}|z_{ij})d\boldsymbol{\theta}d\lambda_{ij}d\lambda_{ij}^* \\
&= \int p(z_{ij}^*|\lambda_{ij}^*, \boldsymbol{\theta})p(\lambda_{ij}^*|\boldsymbol{\theta}, \lambda_{ij}, z_{ij})p(\boldsymbol{\theta}, \lambda_{ij}|z_{ij})d\boldsymbol{\theta}d\lambda_{ij}d\lambda_{ij}^* \\
&= \int p(z_{ij}^*|\lambda_{ij}^*, \boldsymbol{\theta})p(\lambda_{ij}^*|\boldsymbol{\theta})p(\boldsymbol{\theta}, \lambda_{ij}|z_{ij})d\boldsymbol{\theta}d\lambda_{ij}d\lambda_{ij}^*.
\end{aligned} \tag{4.6}$$

However, this integral is analytically intractable. Thus, once there is no co-variable to be observed before a prediction can be made, the Markov chain Monte Carlo simulation approach has the advantage of producing predictions as byproduct of each iteration of the chain. This process will follow the scheme:

1. Extract a sample from the posterior distribution $\boldsymbol{\theta}^{(t)}, {\lambda_{ij}}^{(t)}|z_{ij}$
2. Sample ${\lambda^*_{ij}}^{(t)}|\boldsymbol{\theta}^{(t)}$
3. Sample ${z^*_{ij}}^{(t)}|{\lambda_{ij}}^{(t)}, \boldsymbol{\theta}^{(t)}, {\lambda^*_{ij}}^{(t)}, z_{ij}$.

Consequently, having these samples from the lower triangle it is possible to report statistics associated with this distribution, such as the mean and the median. Finally, in order to execute the inferential process, the methods described in chapter 2 will be used, particularly the algorithms 2.4.1 and 2.4.2, along with the Model Selection Criteria in the next sections.

4.4 Applications

4.4.1 Simulation

Simulations have been implemented to check the convergence efficiency of the algorithms for models described in section 4.2. Artificial run-off triangles, with $n = 20$, were generated for each model. The parameter values were $\mu = 200$, $\sigma^2 = 1$, and $\alpha_i \sim \mathcal{N}(0, 4)$, for $i = 1, \dots, n-1$, $\beta_j \sim \mathcal{N}(0, 16)$ for $j = 1, \dots, n-1$, $\lambda_{ij} \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$ for $i = 1, \dots, n$ and $j = 1, \dots, n-i+1$, and $\nu = 3$.

For the log-anova restrictions were placed in the last parameters, which means that $\alpha_n = -\sum_{i=1}^{n-1} \alpha_i$, and $\beta_n = -\sum_{j=1}^{n-1} \beta_j$. For the log-ancova, only the last restriction applies. For the log-state-space, as already stated in section 4.2, $\alpha_1 = 0$, and $\boldsymbol{\beta}_{i1} = \mathbf{0}$, then in the simulations $\alpha_i \sim \mathcal{N}(\alpha_{i-1}, 4)$, for $i = 2, \dots, n$, and $\beta_{ij} \sim \mathcal{N}(\beta_{i-1j}, 16)$, for $j = 2, \dots, n$ and $i = 1, \dots, n-j+1$.

We have set priors to be non-informative, such that σ^2_μ, σ^2_β, σ^2_α, $\to +\infty$, and a, $b \to 0$. Also, following Gelman et al. (2006), for the uniform prior of σ^2_β, we have let $A \to +\infty$. According to the authors if n is bigger than four the posterior will be proper. Then, as it was stated above, the simulations were run with $n = 20$, therefore there seems to be no reason to worry about the posterior propriety of σ^2_β.

In the case where the errors follow a normal distribution, $50,000$ iterations have been run with a burn-in of $5,000$. Table 4.1 shows simulation results for μ and σ^2, all 95% credible intervals contain the true values.

Table 4.1: Estimation of μ and σ^2 for models assuming normal errors and artificial data simulated from a normal distribution.

	Log-Anova			Log-Ancova			Log-State-Space		
	2.5%	mean	97.5%	2.5%	mean	97.5%	2.5%	mean	97.5%
μ	199.42	200.05	200.65	199.41	199.90	200.39	199.33	200.5	201.66
σ^2	0.61	0.98	1.64	0.89	1.20	1.91	0.55	1.04	1.79

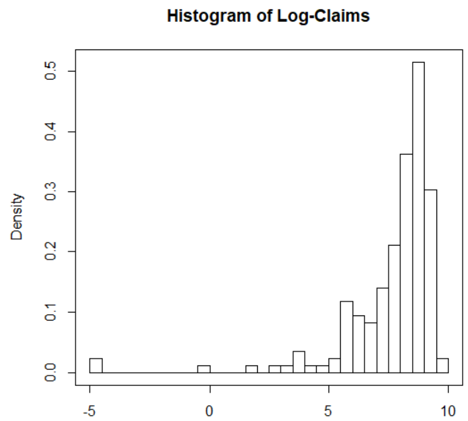


Figure 4.1: Histogram of Log Claims for the data set from Choy et al. (2016).

In the case where the errors follow a student's t distribution, 500,000 iterations were needed, with the first 100,000 being discarded. Table 4.2 displays the estimates for μ, σ^2, and ν for these models. Again, all 95% credible intervals contain the true parameter values.

Table 4.2: Estimation of μ, σ^2 and ν for models assuming student's t errors and artificial data simulated from a student's t distribution.

	Log-Anova			Log-Ancova			Log-State-Space		
	2.5%	mean	97.5%	2.5%	mean	97.5%	2.5%	mean	97.5%
μ	194.91	200.21	205.77	199.45	200.18	202.93	198.34	199.81	201.29
σ^2	0.69	0.94	1.26	0.83	1.11	1.73	0.40	1.13	2.19
ν	2.80	3.37	3.99	2.30	2.77	3.55	2.60	3.12	3.69

4.4.2 Real Data

After simulations, the models were run using the data set from Choy et al. (2016) (see appendix 4.B), its structure can be seen in figure 4.1. The behavior of the claims is very irregular and contains some extreme values. The mean, median, standard deviation, skewness, and excess kurtosis are, respectively, 7.66, 2.05, 8.26, -3.26, 17.53. Also, it is interesting to note its tail behaviors, which might indicate that the estimation of quantiles could be a worthwhile pursuit.

We have used both the Deviance Information Criterion and the Watanabe-Akaike Information Criterion, defined in chapter 2, with the objective of searching for evidence that these data came from a process with a heavy-tailed distribution.

These results are shown in table 4.3. For the normal models 500,000 iterations were run with a burn-in of 100,000. In the student's t cases 1,000,000 iterations were run with a burn-in of 300,000 and thinning of 10.

Table 4.3: Model Selection Criteria for models fitting data from figure 4.1 assuming both normal and student's t errors.

Model		Log-Ancova	Log-Anova	Log-State-Spate
Normal	DIC	591.17	641.61	580.27
	$WAIC$	613.15	645.87	500.74
Student's t	DIC	86.93	155.23	**39.72**
	$WAIC$	268.94	296.06	**127.15**

An example of the predicted values for each model, using the student's t distribution, for the second line of the run-off triangle can be seen in figure 4.2. The log-state-space clearly has the best fist for these data. The residuals for the three models are shown in figure 4.3, they all seem to have asymmetric distributions, with the log-state-space clearly having the smallest variance. Also, table 4.4 shows means and 95% intervals for some parameters of these models.

On one hand, both log-anova and log-ancova have very similar estimates. Which makes sense, once these models are very much alike, with just a slight change in the mean function. On the other hand, the log-state-space produces a result that is excessively different in all three parameters shown in table 4.4. Drawing attention to ν, which is surprisingly smaller. For instance, its 95% interval upper bound, is smaller than the lower bound of this interval for the log-anova.

Table 4.4: Estimation of μ, σ^2, and ν for models fitting data from figure 4.1 assuming student's t errors

	Log-Anova			Log-Ancova			Log-State-Space		
	2.5%	mean	97.5%	2.5%	mean	97.5%	2.5%	mean	97.5%
μ	0.72	6.48	12.18	2.21	6.66	11.72	7.75	8.08	8.29
σ^2	0.07	0.11	0.15	0.05	0.09	0.14	0.003	0.01	0.03
ν	1.26	1.64	2.08	1.09	1.49	2.05	0.63	0.86	1.22

Finally, in order to produce the forecasts for these models, we will be focusing on the reserves for the next year, also known as calendar year (Choy et al., 2016). The position where these reserves are placed in the run-off triangle is shown in table 3.2. Therefore, the next year reserves produced by each model will be represented by the summation of that diagonal. Figure 4.4 brings histograms of those forecasts. Table 4.5 displays quantiles for the next year reserves produced by each model with student's t distribution. The log-state-space has the most concentrated distribution when compared to the other two, which look right-skewed.

Table 4.5: Median and 95% interval of next year reserves estimated from models assuming student's t errors and fitting data from figure 4.1.

Percentiles	2.5%	50%	97.5%
Log-Anova	26,556.94	37,275.89	85,333.61
Log-Ancova	40,516.30	46,613.36	70,746.53
Log-State-Space	29,805.17	44,828.75	58,987.75

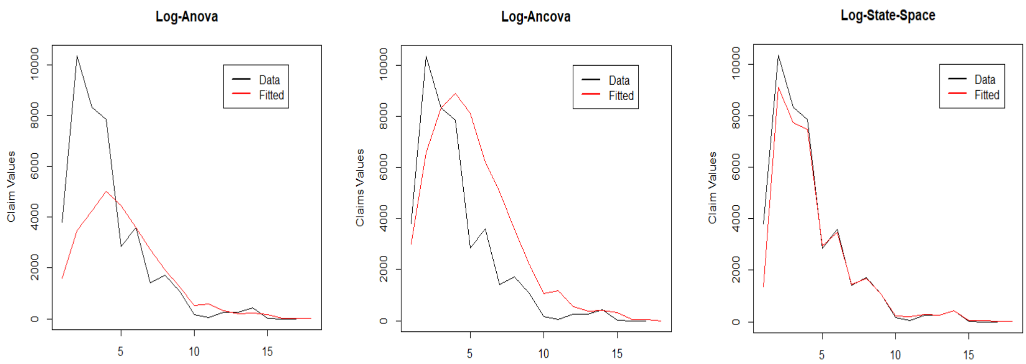
4.5 Discussion

In this chapter models with a focus on estimating future claim mean values were presented. This approach is the most commonly used for modeling reserves.

First of all, simulations were carried out to show that the models could actually recover the true parameter values. Right after that, the results from Choy et al. (2016) have been reproduced using the models described in section 4.2. Strangely, the same results were not obtained. In that paper the authors claimed that the log-ancova was the best model for the data from figure 4.1. However, the results found through the model selection information criteria, suggest that the log-state-space is, in fact, the best model for this data set.

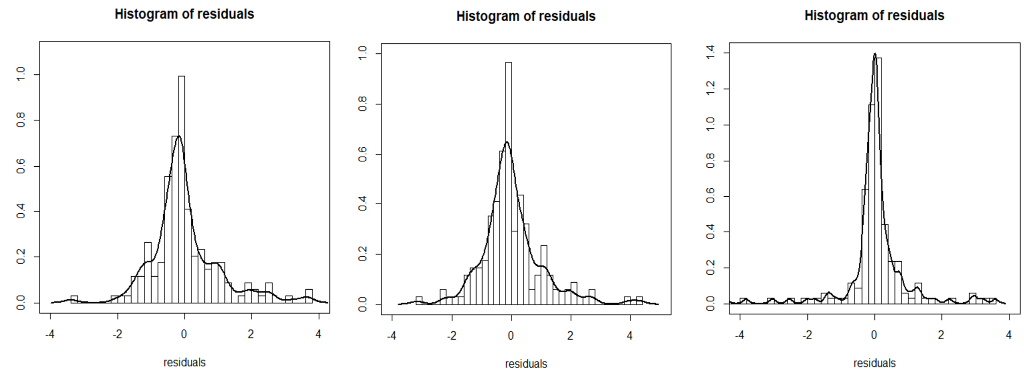
However, when looking at the reserves estimates produced by these models, see table 4.5, it is possible to see that the difference between the central estimate, 50% quantile, for the log-state-space does not differ too much from the one provided by the log-ancova.

In the next chapter we are going to allow the errors to follow an asymmetric Laplace distribution and compare the results with the ones presented in this chapter.



(a) Line 2 Log-Anova (b) Line 2 Log-Ancova (c) Line 2 Log-State-Space

Figure 4.2: Example of fitting the second line of the triangle for each model assuming student's t errors and fitting data from 4.1.



(a) Log-Anova (b) Log-Ancova (c) Log-State-Space

Figure 4.3: Residuals for each model assuming student's t errors and fitting data from figure 4.1.

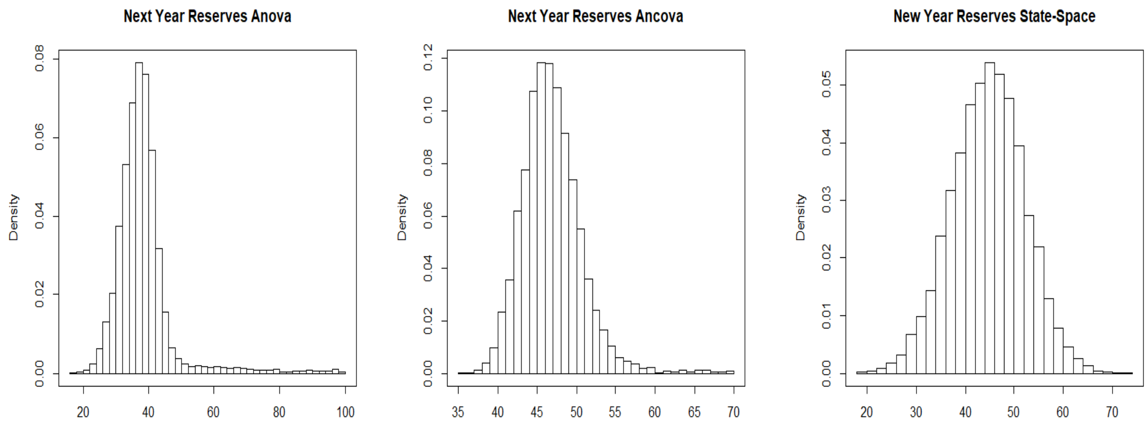


Figure 4.4: Next year reserves for models assuming student's t errors and fiting data from figure 4.1, the values have been divided by 1,000.

Appendix

4.A Full Conditional Posterior Distributions

4.A.1 Log-Anova

This model has the following formulation,

$$
\begin{gathered}
Z_{ij} = \mu_{ij} + \lambda_{ij}{}^{-1/2}\epsilon, \\
\mu_{ij} = \mu + \alpha_i + \beta_j, \\
\sum_{i=1}^{n} \alpha_i = 0 \text{ and } \sum_{j=1}^{n} \beta_j = 0, \\
\lambda_{ij} \sim \Gamma(\frac{\nu}{2}, \frac{\nu}{2}), \\
Z_{ij} \sim \mathcal{N}\left(\mu_{ij},\ \lambda_{ij}{}^{-1}\sigma^2\right),
\end{gathered}
$$

therefore, for μ, considering its prior as $\mu \sim \mathcal{N}(0, \sigma_\mu^2)$

$$
\begin{aligned}
p\left(\mu | \boldsymbol{z}, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \nu\right) &\propto \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \lambda_{ij} \left(z_{ij} - \mu - \alpha_i - \beta_j\right)^2 \right\} \times \exp\left\{ -\frac{\mu^2}{2\sigma_\mu^2} \right\} \\
&\propto \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2} \left(-2z_{ij}\mu + \mu^2 + 2\mu\alpha_i + 2\mu\beta_j\right) \right\} \times \exp\left\{ -\frac{\mu^2}{2\sigma_\mu^2} \right\} \\
&\propto \exp\left\{ -\frac{1}{2}\left[\left(\sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2} + \frac{1}{\sigma_\mu^2} \right)\mu^2 - 2\mu\left(\sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2} \left(z_{ij} - \alpha_i - \beta_j\right) \right) \right] \right\}
\end{aligned}
$$

Completing squares it is possible to see that,

$$
\mu | \boldsymbol{z}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, \nu \ \sim\ \mathcal{N}\left(\frac{\frac{1}{\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \lambda_{ij} \left(z_{ij} - \alpha_i - \beta_j\right)}{\frac{1}{\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \lambda_{ij} + \frac{1}{\sigma_\mu{}^2}}, + \frac{1}{\frac{1}{\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \lambda_{ij} + \frac{1}{\sigma_\mu}^2} \right)
$$

For σ^2, considering its prior as $\sigma^2 \sim \mathcal{IG}\left(a, b\right)$ and $N = \frac{n(n+1)}{2}$

$$p\left(\sigma^2|\boldsymbol{z},\mu,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\lambda},\nu\right) \propto \left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left(\frac{1}{\sigma^2}\right)^{1/2}\exp\left\{-\frac{\lambda_{ij}}{2\sigma^2}(z_{ij}-\mu-\alpha_i-\beta_j)^2\right\}\right\}\times\left(\frac{1}{\sigma^2}\right)^{a+1}\exp\left\{-\frac{b}{\sigma^2}\right\}$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{(N/2\ +\ a)+1}\exp\left\{-\frac{1}{\sigma^2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{2}\left(z_{ij}-\mu-\alpha_i-\beta_j\right)^2+b\right]\right\}$$

Then, it is clear that

$$\sigma^2|\boldsymbol{z},\mu,\boldsymbol{\lambda},\boldsymbol{\alpha},\boldsymbol{\beta},\nu\ \sim \mathcal{IG}\left(\frac{N}{2}+a,\ \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}(z_{ij}-(\mu+\alpha_i+\beta_j))^2\right)+b\right)$$

For a particular α_i, considering its prior as $\alpha_i \sim \mathcal{N}(0,\sigma^2_\alpha)$

$$p\left(\alpha_i|\boldsymbol{z},\mu,\sigma^2,\boldsymbol{\beta},\boldsymbol{\lambda},\nu\right) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n-i+1}\lambda_{ij}\left(z_{ij}-\mu-\alpha_i-\beta_j\right)^2\right\}\times\exp\left\{-\frac{{\alpha_i}^2}{2\sigma^2_\alpha}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{\sigma^2}\left(-2z_{ij}\alpha_i+{\alpha_i}^2+2\alpha_i\mu+2\alpha_i\beta_j\right)\right\}\times\exp\left\{-\frac{{\alpha_i}^2}{2\sigma^2_\alpha}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{\sigma^2}+\frac{1}{\sigma^2_\alpha}\right){\alpha_i}^2-2\alpha_i\left(\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{\sigma^2}\left(z_{ij}-\mu-\beta_j\right)\right)\right]\right\}$$

Once again, completing squares

$$\alpha_i|\boldsymbol{z},\mu,\boldsymbol{\lambda},\boldsymbol{\beta},\sigma^2,\ \nu\ \sim\ \mathcal{N}\left(\frac{\frac{1}{\sigma^2}\sum_{j=1}^{n-i+1}\ \lambda_{ij}\left(z_{ij}-\mu-\beta_j\right)}{\frac{1}{\sigma^2}\sum_{j=1}^{n-i+1}\lambda_{ij}+\frac{1}{{\sigma_\alpha}^2}},\ \frac{1}{\frac{1}{\sigma^2}\sum_{j=1}^{n-i+1}\lambda_{ij}+\frac{1}{{\sigma_\alpha}^2}}\right),$$

for $i=1,\ldots,n-1$.

For a particular β_j, considering its prior as $\beta_j \sim \mathcal{N}(0,\sigma^2_\beta)$

$$p\left(\beta_j|\boldsymbol{z},\mu,\sigma^2,\boldsymbol{\alpha},\boldsymbol{\lambda},\nu\right) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n-j+1}\lambda_{ij}\left(z_{ij}-\mu-\alpha_i-\beta_j\right)^2\right\}\times\exp\left\{-\frac{{\beta_j}^2}{2\sigma^2_\beta}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n-j+1}\frac{\lambda_{ij}}{\sigma^2}\left(-2z_{ij}\beta_j+{\beta_j}^2+2\beta_j\mu+2\beta_j\alpha_i\right)\right\}\times\exp\left\{-\frac{\beta_j 2}{2\sigma^2_\beta}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n-j+1}\frac{\lambda_{ij}}{\sigma^2}+\frac{1}{\sigma^2_\beta}\right){\beta_j}^2-2\beta_j\left(\sum_{i=1}^{n-j+1}\frac{\lambda_{ij}}{\sigma^2}\left(z_{ij}-\mu-\alpha_i\right)\right)\right]\right\}$$

Completing squares

$$\beta_j|\boldsymbol{z},\mu,\boldsymbol{\lambda},\boldsymbol{\alpha},\sigma^2,\nu \sim \mathcal{N}\left(\frac{\frac{1}{\sigma^2}\sum_{i=1}^{n-j+1}\lambda_{ij}\left(z_{ij}-\mu-\alpha_i\right)}{\frac{1}{\sigma^2}\sum_{i=1}^{n-j+1}\lambda_{ij}+\frac{1}{\sigma_\beta{}^2}},\ \frac{1}{\frac{1}{\sigma^2}\sum_{i=1}^{n-j+1}\lambda_{ij}+\frac{1}{\sigma_\beta{}^2}}\right),$$

for $j=1,\ldots,n-1$.

For λ_{ij}, considering its prior as $\lambda_{ij}\sim\Gamma(\frac{\nu}{2},\frac{\nu}{2})$

$$\begin{aligned}
p\left(\lambda_{ij}|z_{ij},\mu,\sigma^2,\alpha_i,\beta_j,\nu\right) &\propto {\lambda_{ij}}^{1/2}\exp\left\{-\frac{\lambda_{ij}}{2\sigma^2}(z_{ij}-\mu-\alpha_i-\beta_j)^2\right\}\times{\lambda_{ij}}^{\nu/2\ -\ 1}\exp\left\{-\lambda_{ij}\frac{\nu}{2}\right\}\\
&\propto {\lambda_{ij}}^{\frac{1+\nu}{2}\ -\ 1}\exp\left\{-\lambda_{ij}\left(\frac{(z_{ij}-\mu-\alpha_i-\beta_j)^2}{2\sigma^2}+\frac{\nu}{2}\right)\right\}
\end{aligned}$$

Then, it is easy to see that

$$\lambda_{ij}|z_{ij},\mu,\alpha_i,\beta_j,\sigma^2,\nu\sim\Gamma\left(\frac{1}{2}+c,\ \frac{1}{2\sigma^2}\ \left(z_{ij}-(\mu+\alpha_i+\beta_j)\right)^2\ +\ d\right),$$

for $i=1,\ldots,n$ and $j=1,\ldots,n-i+1$.

4.A.2 Log-Ancova

The log-ancova has the following formulation,

$$\begin{gathered}
Z_{ij}=\mu_{ij}+{\lambda_{ij}}^{-1/2}\epsilon,\\
\mu_{ij}=\mu+i\alpha+\beta_j,\\
\sum_{j=1}^{n}\beta_j=0,\\
\lambda_{ij}\sim\Gamma(\frac{\nu}{2},\frac{\nu}{2}),\\
Z_{ij}\sim\mathcal{N}\left(\mu_{ij},\ {\lambda_{ij}}^{-1}\sigma^2\right).
\end{gathered}$$

For μ, considering its prior as $\mu\sim\mathcal{N}(0,\sigma_\mu^2)$

$$\begin{aligned}
p\left(\mu|\boldsymbol{z},\sigma^2,\alpha,\boldsymbol{\beta},\boldsymbol{\lambda},\nu\right) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}\left(z_{ij}-\mu-i\alpha-\beta_j\right)^2\right\}\times\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{\sigma^2}\left(-2z_{ij}\mu+\mu^2+2\mu\ i\alpha+2\mu\beta_j\right)\right\}\times\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{\sigma^2}+\frac{1}{\sigma_\mu^2}\right)\mu^2-2\mu\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{\sigma^2}\left(z_{ij}-i\alpha-\beta_j\right)\right)\right]\right\}
\end{aligned}$$

Completing squares it is possible to see that,

$$\mu|\boldsymbol{z},\boldsymbol{\lambda},\alpha,\boldsymbol{\beta},\sigma^2,\nu \sim \ \mathcal{N}\left(\frac{\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}\left(z_{ij}-i\alpha-\beta_j\right)}{\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}+\frac{1}{\sigma_\mu{}^2}}, + \frac{1}{\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}+\frac{1}{\sigma_\mu}{}^2}\right)$$

For σ^2, considering its prior as $\sigma^2 \sim \mathcal{IG}\left(a,b\right)$ and $N=\frac{n(n+1)}{2}$

$$\begin{aligned}
p\left(\sigma^2|\boldsymbol{z},\mu,\alpha,\boldsymbol{\beta},\boldsymbol{\lambda},\nu\right) &\propto \left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left(\frac{1}{\sigma^2}\right)^{1/2}\exp\left\{-\frac{\lambda_{ij}}{2\sigma^2}(z_{ij}-\mu-i\alpha-\beta_j)^2\right\}\right\}\times\left(\frac{1}{\sigma^2}\right)^{a+1}\exp\left\{-\frac{b}{\sigma^2}\right\}\\
&\propto \left(\frac{1}{\sigma^2}\right)^{(N/2\ +\ a)+1}\exp\left\{-\frac{1}{\sigma^2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{2}\left(z_{ij}-\mu-i\alpha-\beta_j\right)^2+b\right]\right\}
\end{aligned}$$

Then, it is clear that

$$\sigma^2|\boldsymbol{z},\mu,\boldsymbol{\lambda},\alpha,\boldsymbol{\beta},\nu \sim \mathcal{IG}\left(\frac{N}{2}+a,\ \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}(Z_{ij}-(\mu+i\alpha+\beta_j))^2\right)+b\right)$$

For α, considering its prior as $\alpha \sim \mathcal{N}(0,\sigma^2_\alpha)$

$$\begin{aligned}
p\left(\alpha|\boldsymbol{z},\mu,\sigma^2,\boldsymbol{\beta},\boldsymbol{\lambda},\nu\right) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}\left(z_{ij}-\mu-i\alpha-\beta_j\right)^2\right\}\times\exp\left\{-\frac{\alpha^2}{2\sigma^2_\alpha}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{\sigma^2}\left(-2z_{ij}\ i\alpha+(i\alpha)^2+2i\alpha\mu+2i\alpha\beta_j\right)\right\}\times\exp\left\{-\frac{\alpha 2}{2\sigma^2_\alpha}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{i^2\ \lambda_{ij}}{\sigma^2}+\frac{1}{\sigma^2_\alpha}\right)\alpha^2-2\alpha\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{i\ \lambda_{ij}}{\sigma^2}\left(z_{ij}-\mu-\beta_j\right)\right)\right]\right\}
\end{aligned}$$

Once again, completing squares

$$\alpha|\ \boldsymbol{z},\ \mu,\ \boldsymbol{\lambda},\ \boldsymbol{\beta},\ \sigma^2,\ \nu\ \sim\ \mathcal{N}\left(\frac{\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\ i\ \lambda_{ij}\left(z_{ij}-\mu-\beta_j\right)}{\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}i^2\ \lambda_{ij}+\frac{1}{\sigma_\alpha{}^2}},\frac{1}{\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}i^2\ \lambda_{ij}+\frac{1}{\sigma_\alpha{}^2}}\right).$$

For a particular β_j, considering its prior as $\beta_j \sim \mathcal{N}(0,\sigma^2_\beta)$

$$\begin{aligned}
p\left(\beta_j|\boldsymbol{z},\mu,\sigma^2,\alpha,\boldsymbol{\lambda},\nu\right) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n-j+1}\lambda_{ij}\left(z_{ij}-\mu-i\alpha-\beta_j\right)^2\right\}\times\exp\left\{-\frac{\beta_j{}^2}{2\sigma^2_\beta}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n-j+1}\frac{\lambda_{ij}}{\sigma^2}\left(-2z_{ij}\beta_j+\beta_j{}^2+2\beta_j\mu+2\beta_j\ i\alpha\right)\right\}\times\exp\left\{-\frac{\beta_j 2}{2\sigma^2_\beta}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n-j+1}\frac{\lambda_{ij}}{\sigma^2}+\frac{1}{\sigma^2_\beta}\right)\beta_j{}^2-2\beta_j\left(\sum_{i=1}^{n-j+1}\frac{\lambda_{ij}}{\sigma^2}\left(z_{ij}-\mu-i\alpha\right)\right)\right]\right\}
\end{aligned}$$

Completing squares

$$\beta_j|\boldsymbol{z},\mu,\boldsymbol{\lambda},\boldsymbol{\alpha},\sigma^2,\nu \sim \ \mathcal{N}\left(\frac{\frac{1}{\sigma^2}\sum_{i=1}^{n-j+1}\ \lambda_{ij}\left(z_{ij}-\mu-i\alpha\right)}{\frac{1}{\sigma^2}\sum_{i=1}^{n-j+1}\ \lambda_{ij}+\frac{1}{\sigma_\beta{}^2}},\ \frac{1}{\frac{1}{\sigma^2}\sum_{i=1}^{n-j+1}\ \lambda_{ij}+\frac{1}{\sigma_\beta{}^2}}\right),$$

for $j = 1, \ldots, n-1$.

For λ_{ij}, considering its prior as $\lambda_{ij} \sim \Gamma(\frac{\nu}{2},\frac{\nu}{2})$

$$\begin{aligned}
p\left(\lambda_{ij}|z_{ij},\mu,\sigma^2,\alpha,\beta_j,\nu\right) &\propto \lambda_{ij}{}^{1/2}\exp\left\{-\frac{\lambda_{ij}}{2\sigma^2}(z_{ij}-\mu-i\alpha-\beta_j)^2\right\}\times\lambda_{ij}{}^{\nu/2\ -\ 1}\exp\left\{-\lambda_{ij}\frac{\nu}{2}\right\}\\
&\propto \lambda_{ij}{}^{\frac{1+\nu}{2}\ -\ 1}\exp\left\{-\lambda_{ij}\left(\frac{(z_{ij}-\mu-i\alpha-\beta_j)^2}{2\sigma^2}+\frac{\nu}{2}\right)\right\}
\end{aligned}$$

Then, it is easy to see that

$$\lambda_{ij}|z_{ij},\mu,\alpha,\beta_j,\sigma^2,\nu \sim \Gamma\left(\frac{1}{2}+c,\ \frac{1}{2\sigma^2}\ \left(z_{ij}-(\mu+i\alpha+\beta_j)\right)^2\ +\ d\right),$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, n-i+1$.

4.A.3 Log-State-Space

The log-state=space has the following formulation,

$$\begin{gathered}
Z_{ij} = \mu_{ij} + \lambda_{ij}{}^{-1/2}\epsilon,\\
\mu_{ij} = \mu + \alpha_i + \beta_{ij},\\
\alpha_1 = 0 \text{ and } \boldsymbol{\beta_{i1}} = \boldsymbol{0},\\
\lambda_{ij} \sim \Gamma(\frac{\nu}{2},\frac{\nu}{2}),\\
Z_{ij} \sim \mathcal{N}\left(\mu_{ij},\ \lambda_{ij}{}^{-1}\sigma^2\right).
\end{gathered}$$

For μ, considering its prior as $\mu \sim \mathcal{N}(0,\sigma_\mu^2)$

$$\begin{aligned}
p\left(\mu|\boldsymbol{z},\sigma^2,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\lambda},\nu\right) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}\left(z_{ij}-\mu-\alpha_i-\beta_{ij}\right)^2\right\}\times\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{\sigma^2}\left(-2z_{ij}\mu+\mu^2+2\mu\alpha_i+2\mu\beta_{ij}\right)\right\}\times\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{\sigma^2}+\frac{1}{\sigma_\mu^2}\right)\mu^2-2\mu\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\lambda_{ij}}{\sigma^2}\left(z_{ij}-\alpha_i-\beta_{ij}\right)\right)\right]\right\}
\end{aligned}$$

Completing squares it is possible to see that,

$$\mu|\boldsymbol{z},\boldsymbol{\lambda},\boldsymbol{\alpha},\boldsymbol{\beta},\sigma^2,\nu \sim \mathcal{N}\left(\frac{\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}\left(z_{ij}-\alpha_i-\beta_{ij}\right)}{\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}+\frac{1}{\sigma_\mu{}^2}},+\frac{1}{\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}+\frac{1}{\sigma_\mu}^2}\right)$$

For σ^2, considering its prior as $\sigma^2 \sim \mathcal{IG}\left(a, b\right)$ and $N = \frac{n(n+1)}{2}$

$$p\left(\sigma^2 | \boldsymbol{z}, \mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \nu\right) \propto \left\{ \prod_{i=1}^{n} \prod_{j=1}^{n-i+1} \left(\frac{1}{\sigma^2}\right)^{1/2} \exp\left\{ -\frac{\lambda_{ij}}{2\sigma^2}\left(z_{ij} - \mu - \alpha_i - \beta_{ij}\right)^2 \right\} \right\} \times \left(\frac{1}{\sigma^2}\right)^{a+1} \exp\left\{ -\frac{b}{\sigma^2} \right\}$$
$$\propto \left(\frac{1}{\sigma^2}\right)^{(N/2\ +\ a)+1} \exp\left\{ -\frac{1}{\sigma^2}\left[\sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{2} \left(z_{ij} - \mu - \alpha_i - \beta_{ij}\right)^2 + b \right] \right\}$$

Then, it is clear that

$$\sigma^2 | \boldsymbol{z}, \mu, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \nu \sim \mathcal{IG}\left(\frac{N}{2} + a,\ \frac{1}{2}\left(\sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \lambda_{ij}(z_{ij} - (\mu + \alpha_i + \beta_{ij}))^2 \right) + b \right)$$

For a particular α_i, considering its prior as $\alpha_i \sim \mathcal{N}(\alpha_{i-1}, \sigma_\alpha^2)$

$$p\left(\alpha_i | \boldsymbol{z}, \mu, \sigma^2, \boldsymbol{\beta}, \boldsymbol{\lambda}, \nu\right) \propto \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{n-i+1} \lambda_{ij}\left(z_{ij} - \mu - \alpha_i - \beta_{ij}\right)^2 \right\} \times \exp\left\{ -\frac{\left(\alpha_i - \alpha_{i-1}\right)^2}{2\sigma_\alpha^2} \right\}$$
$$\times \exp\left\{ -\frac{\left(\alpha_{i+1} - \alpha_i\right)^2}{\sigma_\alpha^2} \right\}$$
$$\propto \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2}\left(-2z_{ij}\alpha_i + {\alpha_i}^2 + 2\alpha_i\mu + 2\alpha_i\beta_{ij}\right) \right\} \times \exp\left\{ -\frac{\left({\alpha_i}^2 - 2\alpha_{i-1}\alpha_i\right)}{2\sigma_\alpha^2} \right\}$$
$$\times \exp\left\{ -\frac{\left(-2\alpha_{i+1}\alpha_i + {\alpha_i}^2\right)}{2\sigma_\alpha^2} \right\}$$
$$\propto \exp\left\{ -\frac{1}{2}\left[\left(\sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2} + \frac{2}{\sigma_\alpha^2} \right){\alpha_i}^2 - 2\alpha_i \left(\frac{\alpha_{i-1} + \alpha_{i+1}}{\sigma_\alpha^2} + \sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2}\left(z_{ij} - \mu - \beta_{ij}\right) \right) \right] \right\}$$

Once again, completing squares

$$\alpha_i | \mathbf{z}, \mu, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \nu \sim \mathcal{N}\left(\frac{\frac{\alpha_{i-1}+\alpha_{i+1}}{\sigma_\alpha^2} + \sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2}\left(z_{ij} - \mu - \beta_{ij}\right)}{\sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2} + \frac{2}{\sigma_\alpha^2}}, \frac{1}{\sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2} + \frac{2}{\sigma_\alpha^2}} \right),$$

for $i = 2, \ldots, n-1$.

It is also straightforward to see that for $i = n$

$$\alpha_i | \mathbf{z}, \mu, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \nu \sim \mathcal{N}\left(\frac{\frac{\alpha_{i-1}}{\sigma_\alpha^2} + \sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2}\left(z_{ij} - \mu - \beta_{ij}\right)}{\sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2} + \frac{1}{\sigma_\alpha^2}}, \frac{1}{\sum_{j=1}^{n-i+1} \frac{\lambda_{ij}}{\sigma^2} + \frac{1}{\sigma_\alpha^2}} \right),$$

for$i = n$.

For a particular β_{ij}, considering its prior as $\beta_{ij} \sim \mathcal{N}(\beta_{i-1j}, \sigma_\beta^2)$

$$
\begin{aligned}
p\left(\beta_j|z_{ij},\mu,\sigma^2,\alpha_i,\lambda_{ij},\nu\right) &\propto \exp\left\{-\frac{1}{2\sigma^2}\lambda_{ij}\left(z_{ij}-\mu-\alpha_i-\beta_{ij}\right)^2\right\}\times\exp\left\{-\frac{\left(\beta_{ij}-\beta_{i-1j}\right)^2}{2\sigma^2_\beta}\right\}\\
&\times\exp\left\{-\frac{\left(\beta_{i+1j}-\beta_{ij}\right)^2}{2\sigma^2_\beta}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\frac{\lambda_{ij}}{\sigma^2}\left(-2z_{ij}\beta_{ij}+{\beta_{ij}}^2+2\beta_{ij}\mu+2\beta_{ij}\alpha_i\right)\right\}\times\exp\left\{-\frac{\left({\beta_{ij}}^2-2\beta_{ij}\beta_{i-j}\right)}{2\sigma^2_\beta}\right\}\\
&\times\exp\left\{-\frac{\left(-2\beta_{i+1j}\beta_{ij}+{\beta_{ij}}^2\right)}{2\sigma^2_\beta}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{\lambda_{ij}}{\sigma^2}+\frac{2}{\sigma^2_\beta}\right){\beta_{ij}}^2-2\beta_j\left(\frac{\left(\beta_{i+1j}+\beta_{i-1j}\right)}{\sigma^2_\beta}+\frac{\lambda_{ij}}{\sigma^2}\left(z_{ij}-\mu-\alpha_i\right)\right)\right]\right\}
\end{aligned}
$$

Completing squares

$$
\beta_{ij}|z_{ij},\mu,\alpha_i,\sigma^2,\lambda_{ij},\nu \sim \mathcal{N}\left(\frac{\frac{\beta_{i-1j}+\beta_{i+1j}}{\sigma^2_\beta}+\frac{\lambda_{ij}}{\sigma^2}\left(z_{ij}-\mu-\alpha_i\right)}{\frac{\lambda_{ij}}{\sigma^2}+\frac{2}{\sigma^2_\beta}},\frac{1}{\frac{\lambda_{ij}}{\sigma^2}+\frac{2}{\sigma^2_\beta}}\right),
$$

for $i=1,\ldots,n-j$ and $j=2,\ldots,n$.

Again, it is straightforward to see that

$$
\beta_{ij}|z_{ij},\mu,\alpha_i,\sigma^2,\lambda_{ij},\nu \sim \mathcal{N}\left(\frac{\frac{\beta_{i-1j}}{\sigma^2_\beta}+\frac{\lambda_{ij}}{\sigma^2}\left(z_{ij}-\mu-\alpha_i\right)}{\frac{\lambda_{ij}}{\sigma^2}+\frac{1}{\sigma^2_\beta}},\frac{1}{\frac{\lambda_{ij}}{\sigma^2}+\frac{1}{\sigma^2_\beta}}\right),
$$

for $i=n-j+1$ and $j=2,\ldots,n$.

For λ_{ij}, considering its prior as $\lambda_{ij}\sim\Gamma(\frac{\nu}{2},\frac{\nu}{2})$

$$
\begin{aligned}
p\left(\lambda_{ij}|z_{ij},\mu,\sigma^2,\alpha_i,\beta_j,\nu\right) &\propto {\lambda_{ij}}^{1/2}\exp\left\{-\frac{\lambda_{ij}}{2\sigma^2}(z_{ij}-\mu-\alpha_i-\beta_{ij})^2\right\}\times{\lambda_{ij}}^{\nu/2\;-\;1}\exp\left\{-\lambda_{ij}\frac{\nu}{2}\right\}\\
&\propto {\lambda_{ij}}^{\frac{1+\nu}{2}\;-\;1}\exp\left\{-\lambda_{ij}\left(\frac{\left(z_{ij}-\mu-\alpha_i-\beta_{ij}\right)^2}{2\sigma^2}+\frac{\nu}{2}\right)\right\}
\end{aligned}
$$

Then, it is easy to see that

$$
\lambda_{ij}\mid z_{ij},\ \mu,\ \alpha_i,\ \beta_{ij},\ \sigma^2,\ \nu\ \sim\Gamma\left(\frac{1}{2}+c,\ \frac{1}{2\sigma^2}\ \left(z_{ij}-(\mu+\alpha_i+\beta_{ij})\right)^2\ +\ d\right),
$$

for$i=1,\ldots,n$ and $j=1,\ldots,n-i+1$.

The variance of $\boldsymbol{\beta}$ will receive a uniform prior $\sigma^2_\beta\sim\mathcal{U}(0,A)$, with $A\to\infty$

$$
\begin{aligned}
p\left(\sigma^2_\beta|\boldsymbol{\beta}\right) &\propto \prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left(\frac{1}{\sigma^2_\beta}\right)^{1/2}\exp\left\{-\frac{\left(\beta_{ij}-\beta_{i-1j}\right)^2}{2\sigma^2_\beta}\right\}\\
&\propto \left(\frac{1}{\sigma^2_\beta}\right)^{N/2}\exp\left\{-\frac{1}{\sigma^2_\beta}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\left(\beta_{ij}-\beta_{i-1j}\right)^2}{2}\right\}
\end{aligned}
$$

Therefore,

$$\sigma_\beta^2|\boldsymbol{\beta} \sim \mathcal{IG}\left(\frac{N}{2}+1, \sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{(\beta_{ij}-\beta_{i-1j})^2}{2}\right)$$

Finally, for all models above, the posterior distribution of ν is obtained by simple multiplication, once it does not have a closed analytical form

$$p\left(\nu \mid \boldsymbol{z},\ \mu,\ \boldsymbol{\lambda},\ \boldsymbol{\alpha},\ \boldsymbol{\beta},\ \sigma^2\ \right) \propto \frac{\left(\frac{\nu}{2}\right)^{\frac{N\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)^N}\ \exp\left\{-\frac{\nu}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\lambda_{ij}\right\}\left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\lambda_{ij}^{\frac{\nu}{2}-1}\right\}\times$$
$$\times\left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}}\left[\psi'\left(\frac{\nu}{2}\right)-\psi'\left(\frac{\nu+1}{2}\right)-\frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}.$$

4.B Claim Amounts

Table 4.6: Claim amounts paid to insureds of an insurance product form 1978 to 1995 (Choy et al., 2016).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1978	3323	8332	9572	10172	7631	3855	3252	4433	2188	333	199	692	311	0	405	293	76	14
1979	3785	10342	8330	7849	2839	3577	1404	1721	1065	156	35	259	250	420	6	1	0	
1980	4677	9989	8745	10228	8572	5787	3855	1445	1612	626	1172	589	438	473	370	31		
1981	5288	8089	12839	11828	7560	6383	4118	3016	1575	1985	2645	266	38	45	115			
1982	2294	9869	10242	13808	8775	5419	2424	1597	4149	1296	917	295	428	359				
1983	3600	7514	8247	9327	8584	4245	4096	3216	2014	593	1188	691	368					
1984	3642	7394	9838	9733	6377	4884	11920	4188	4492	1760	944	921						
1985	2463	5033	6980	7722	6702	7834	5579	3622	1300	3069	1370							
1986	2267	5959	6175	7051	8102	6339	6978	4396	3107	903								
1987	2009	3700	5298	6885	6477	7570	5855	5751	3871									
1988	1860	5282	3640	7538	5157	5766	6862	2572										
1989	2331	3517	5310	6066	10149	9265	5262											
1990	2314	4487	4112	7000	11163	10057												
1991	2607	3952	8228	7895	9317													
1992	2595	5403	6579	15546														
1993	3155	4974	7961															
1994	2626	5704																
1995	2827																	

Chapter 5

Models with Asymmetric Laplace Distribution

5.1 Introduction

In this chapter, the class of models using the asymmetric Laplace distribution is presented. These models will be used in the quantile estimation of reserves, as opposed to the last chapter where the mean estimation was the main goal. Section 5.2 brings the formulation for each model. Section 5.3 brings a concise discussion of the Bayesian inferential process for these models. Section 5.4 presents results for a study with artificial data sets, where it is shown how the asymmetric Laplace can reasonably recover the quantiles of a process regardless the distribution that generated it. This section also displays applications with real-world data for these models, comparing them to the ones presented in the previous chapter. Finally, this chapter ends with a discussion in section 5.5.

5.2 Models with ALD

This section will bring the formulations for the models already described in section 4.2 using the asymmetric Laplace distribution. The models descriptions will be shorter, once the basic ideas behind each model do not change from the last chapter.

5.2.1 Log-Anova

In the loss reserving case, once again considering Y_{ij} as being the amount paid to insureds from policy-year i with $j-1$ years of delay and $Z_{ij} = log(Y_{ij})$, the log-anova model will have the following formulation, immediately derived from equation 3.14, for all $j \leq n-i+1,\ i \leq n$,

$$\begin{aligned} Z_{ij} &= \mu_{ij} + \theta\nu_{ij} + \tau\sqrt{\sigma\nu_{ij}}u_{ij}, \\ \mu_{ij} &= \mu + \alpha_i + \beta_j, \\ u_{ij} &\sim \mathcal{N}(0,1), \\ \nu_{ij} &\sim \mathcal{E}\left(\sigma\right), \end{aligned} \tag{5.1}$$

where, $\mathcal{N}(.,.)$ and $\mathcal{E}(.)$ are normal and exponential distributions, respectively. Also, following this representation, according to proposition 1, if we let

$$\epsilon_{ij} = \theta\nu_{ij} + \tau\sqrt{\sigma\nu_{ij}}u_{ij}, \tag{5.2}$$

then, $\epsilon_{ij} \sim ALD\,(0, \sigma, p)$, where $ALD(.,.,.)$ is an asymmetric Laplace distribution.

According to this representation $Z_{ij} \sim \mathcal{N}\left(\mu_{ij} + \theta\nu_{ij}, \tau^2\sigma\nu_{ij}\right)$, with θ and τ are defined in equation 3.11. The assumptions will continue to be the same as the ones described in section 3.5, with restriction $\sum_i \alpha_i = \sum_j \beta_j = 0$.

5.2.2 Log-Ancova

Using again the index of the policy year i as a regressor, the log-ancova will have the same formulation as the one stated in equation 5.1, for all $j \leq n - i + 1,\ i \leq n$ except for the mean function which will be given by

$$\mu_{ij} = \mu + i\alpha + \beta_j, \tag{5.3}$$

and the assumptions will be the ones states in section 3.5, with the only restriction $\sum_j \beta_j = 0$.

5.2.3 Log-State-Space

The state-space model will have the formulation stated in 5.1, with mean function given by

$$\mu_{ij} = \mu + \alpha_i + \beta_{ij}, \tag{5.4}$$

with α and β following random walks

$$\begin{aligned} \alpha_i &= \alpha_{i-1} + \epsilon_\alpha, \\ \beta_{ij} &= \beta_{i-1j} + \epsilon_\beta, \end{aligned} \tag{5.5}$$

where $\epsilon_\alpha \sim \mathcal{N}(0, \sigma^2_\alpha)$, and $\epsilon_\beta \sim \mathcal{N}(0, \sigma^2_\beta)$. Assumptions are the same as the ones described in section 3.6.

5.3 Bayesian Inference

Again, let $\boldsymbol{\theta}$ be the vector with all the parameters and $\boldsymbol{z}$ be the vector with all observations, then the likelihood function in this case can be written as

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{z}) = \prod_{i=1}^{n} \prod_{j=1}^{n-i+1} \frac{1}{\sqrt{2\pi\tau^2\sigma\nu_{ij}}} \exp\left\{ -\frac{1}{2\tau^2\sigma\nu_{ij}}(z_{ij} - \mu_{ij} - \theta\nu_{ij})^2 \right\} I(\nu_{ij} > 0), \tag{5.6}$$

where μ_{ij} is going to be the mean function for each model described in section 5.2.

5.3.1 Prior Distributions

For the parameters μ, α_i, for all i, β_j, forall j the priors stated in section 4.3 will be used here, the only exceptions are the parameters ν_{ij} and σ. These will receive the following prior distributions: $\nu_{ij} \sim \mathcal{E}(\sigma)$ and $\sigma \sim \mathcal{IG}(a, b)$. Where $\mathcal{E}(.)$ is an exponential distribution and $\mathcal{IG}(.,.)$ is an inverse-gamma distribution.

5.3.2 The Joint Posterior Distribution

Once again, as in section 4.3, the joint posterior distribution will be obtained by the means discussed in chapter 2, especially by equation 2.2. Therefore, the joint posterior distribution of log-anova will be:

$$
\begin{aligned}
p\left(\mu, \sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\nu} | \boldsymbol{z}\right) \propto & \left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left(\frac{1}{\sigma\nu_{ij}}\right)^{1/2}\exp\left\{-\frac{1}{2\tau^2\sigma\nu_{ij}}(z_{ij}-\mu-\alpha_i-\beta_j-\theta\nu_{ij})^2\right\}I(\nu_{ij}>0)\right\}\times \\
& \left\{\prod_{i=1}^{n-1}\exp\left\{-\frac{\alpha_i^2}{2\sigma_\alpha^2}\right\}\right\}\times\left\{\prod_{j=1}^{n-1}\exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\right\}\times\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\}\times\left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\frac{1}{\sigma}\exp\left\{-\frac{\nu_{ij}}{\sigma}\right\}I(\nu_{ij}>0)\right\}\times \\
& \left\{\left(\frac{1}{\sigma^2}\right)^{a+1}\exp\left\{-\frac{b}{\sigma^2}\right\}\right\}.
\end{aligned}
$$

The only difference from this expression for the log-ancova is the slight change in the mean function and the prior for the single α parameter

$$
\exp\left\{-\frac{\alpha^2}{2\sigma_\alpha^2}\right\}.
$$

Lastly, for the log-state-space there will be the changes already seen in chapter 4, where the mean function and the prior distributions for α and β suffer modifications. The joint posterior distribution will be given by:

$$
\begin{aligned}
p\left(\mu, \sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\nu}, \sigma_\beta^2 | \boldsymbol{z}\right) \propto & \left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left(\frac{1}{\sigma\nu_{ij}}\right)^{1/2}\exp\left\{-\frac{1}{2\tau^2\sigma\nu_{ij}}(z_{ij}-\mu-\alpha_i-\beta_{ij}-\theta\nu_{ij})^2\right\}I(\nu_{ij}>0)\right\}\times \\
& \left\{\prod_{i=2}^{n}\exp\left\{-\frac{(\alpha_i-\alpha_{i-1})^2}{2\sigma_\alpha^2}\right\}\right\}\times\left\{\prod_{j=2}^{n}\prod_{i=1}^{n-j+1}\exp\left\{-\frac{(\beta_{ij}-\beta_{i-1j})^2}{2\sigma_\beta^2}\right\}\right\}\times\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\}\times \\
& \left\{\left(\frac{1}{\sigma^2}\right)^{a+1}\exp\left\{-\frac{b}{\sigma^2}\right\}\right\}\times\left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\frac{1}{\sigma}\exp\left\{-\frac{\nu_{ij}}{\sigma}\right\}I(\nu_{ij}>0)\right\}.
\end{aligned}
$$

Once again, as in the last chapter, these expressions do not have closed forms. Therefore, Markov chain Monte Carlo methods will be used in the parameter estimation process. In this case, following Kozumi and Kobayashi (2011), all full conditional posterior distributions will have closed forms, thus we will only need the Gibbs algorithm. The sampling process works as follows:

Gibbs sampling

1. At $t=0$ set initial values $\boldsymbol{\theta}^{(0)} = \{\mu^{(0)}\sigma^{(0)}, \boldsymbol{\alpha}^{(0)}\boldsymbol{\beta}^{(0)}, \boldsymbol{\nu}^{(0)}\}$;

2. Obtain new values $\boldsymbol{\theta}^{(t)} = \{\mu^{(t)}\sigma^{(t)}, \boldsymbol{\alpha}^{(t)}\boldsymbol{\beta}^{(t)}, \boldsymbol{\nu}^{(t)}\}$ by randomly sampling from:

$$\mu^{(t)} \sim p\left(\mu|\sigma^{(t-1)}, \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \boldsymbol{\nu}^{(t-1)}, \boldsymbol{z}\right)$$
$$\sigma^{(t)} \sim p\left(\sigma|\mu^{(t)}, \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \boldsymbol{\nu}^{(t-1)}, \boldsymbol{z}\right)$$
$$\boldsymbol{\alpha}^{(t)} \sim p\left(\boldsymbol{\alpha}|\mu^{(t)}, \sigma^{(t)}, \boldsymbol{\beta}^{(t-1)}, \boldsymbol{\nu}^{(t-1)}, \boldsymbol{z}\right)$$
$$\boldsymbol{\beta}^{(t)} \sim p\left(\boldsymbol{\beta}|\mu^{(t)}, \sigma^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\nu}^{(t-1)}, \boldsymbol{z}\right)$$
$$\boldsymbol{\nu}^{(t)} \sim p\left(\boldsymbol{\nu}|\mu^{(t)}, \sigma^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t)}, \boldsymbol{z}\right)$$

3. Repeat the previous step until convergence is reached,

where $p(\theta_i|\boldsymbol{\theta}_{-i}, \boldsymbol{z})$ is the full conditional posterior distribution of parameter θ_i. A derivation of those distributions is available in appendix $5.A$. Also, in the log-state-space an additional step will be included in order to estimate σ^2_β and in the log-ancova α will be a single scalar parameter.

5.3.3 Predictions

The same reasoning applied to generate the predictions in the previous chapter will be used once more. The target of this procedure is the predictive distribution $p\left(z^*_{ij}|z_{ij}, p\right)$, where z^*_{ij} is the log of unobserved claims, z_{ij} is the data available, and p is the skew parameter of the asymmetric Laplace distribution, $\forall i, j$. Let $\boldsymbol{\theta}$ be a vector with all the parameters, excluding the latent variable ν_{ij}, and let ν^*_{ij} be associated with the unobserved z^*_{ij}. Then, the predictive would be

$$p(z^*_{ij}|z_{ij}, p) = \int p(z^*_{ij}|\nu^*_{ij}, \boldsymbol{\theta}, p)p(\nu^*_{ij}|\boldsymbol{\theta}, p)p(\boldsymbol{\theta}, \nu_{ij}|z_{ij}, p)d\boldsymbol{\theta}d\nu_{ij}d\nu^*_{ij}. \tag{5.7}$$

Therefore, following an iterative process at each MCMC iteration, the predictions will be generated according to the following procedure:

1. Extract a sample from the posterior distribution $\boldsymbol{\theta}^{(t)}, {\nu_{ij}}^{(t)}|z_{ij}, p$

2. Sample ${\nu^*_{ij}}^{(t)}|\boldsymbol{\theta}^{(t)}, p$

3. Sample ${z^*_{ij}}^{(t)}|{\lambda_{ij}}^{(t)}, \boldsymbol{\theta}^{(t)}, {\lambda^*_{ij}}^{(t)}, z_{ij}, p.$

5.4 Applications

5.4.1 Study with Artificial Datasets

In this section the main objective is going to be showing that the asymmetric Laplace can recover the quantiles of the underlying distribution of a process, then this idea is going to be shown in the loss reserving context.

Thus, first of all, consider again the single parameter model of equation 3.8. Generating 100 random values of a student's t distribution with $\nu = 10$ and $\mu = 5$, figure 5.1 show the results of this simulation, for p equals 0.05, 0.5 and 0.95, giving a prior $\mu_p \sim \mathcal{N}(0, \sigma_\mu^2)$, with $\sigma_\mu^2 \to \infty$. The red lines represent the estimated mean values and the green lines are the true quantiles of y_t, both given in table 5.1. This table also brings 95% intervals and standard errors of those estimations.

Table 5.1: Quantile Estimation of single parameter model for data simulated from a student's t and a model fitted assuming asymmetric Laplace errors.

Quantiles of y_t				
p	y_t	Mean of μ_p	95% Interval	S.E.
0.05	3.07	3.04	(2.87; 3.22)	0.09
0.50	5.09	5.10	(4.95; 5.26)	0.12
0.95	6.90	6.91	(6.77; 7.10)	0.07

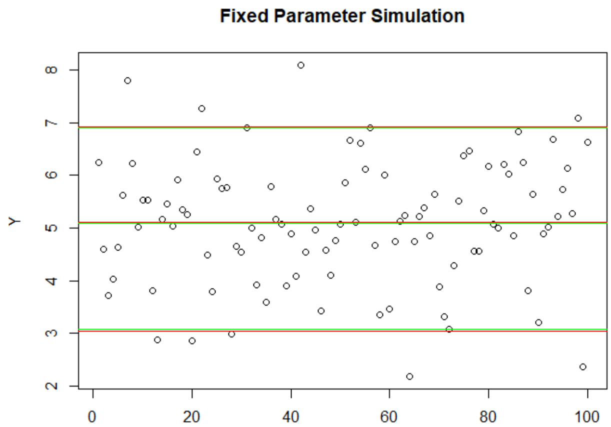
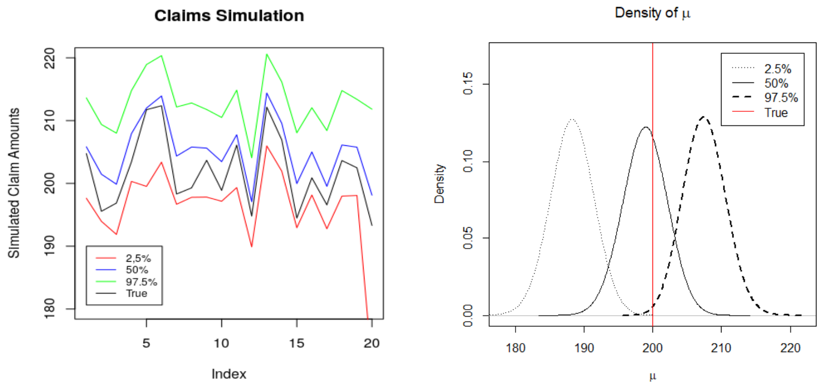


Figure 5.1: Estimation of single parameter model for data simulated from a student's t and a model fitted assuming asymmetric Laplace errors. Red lines are the estimated quantiles and green lines are the true simulated quantiles of y_t.

In the loss reserving context, the idea was to simulate claims from normal and student's t distributions, and fit models with asymmetric Laplace in order to check if the latter could recover the true simulated values. The artificial parameter values applied here have been the same as the ones used to generate the simulations in chapter 4.

Figure 5.2 *a* presents the result for the log-state-space with normally distributed errors for the first line of the triangle, the colored lines are the mean values of each quantile fitted with asymmetric Laplace distribution. Figure 5.2 *b* shows the estimation of μ for the same model, it is possible to see that the median, $p = 0.5$, was able to recover the true simulated value of μ. In these cases 500,000 iterations were run, with the first 100,000 being discarded. These results show the flexibility of the quantile estimation and, more importantly, they corroborate the initial



(a) Simulated Claim Amounts (b) μ

Figure 5.2: Figures a shows the fitting of the quantile version of the state-space with simulated claim values for the first line of the triangle, the colored lines represent the mean of each value, and figure b displays the estimation of the μ parameter for the same model with normally distributed errors.

idea that independently of the underlying distribution of a particular phenomenon it is possible to reasonably recover its quantiles using the ALD.

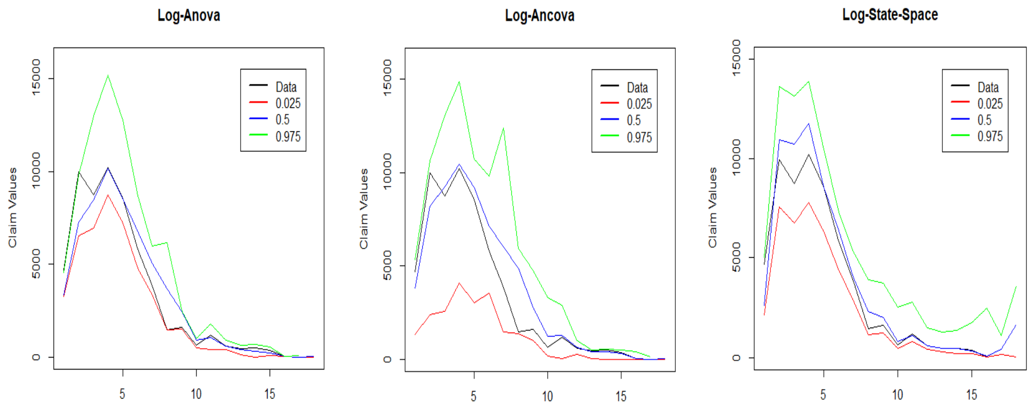
5.4.2 Real Data

The data from Choy et al. (2016) will be used one more time. Figure 4.1 from chapter 4 shows its structure. Our goal now is to estimate the quantiles of the claims distribution and compare the results with ones obtained by the mean models in the previous chapter, specifically the log-state-space, since evidence provided by information criteria explicitly tells us that this is the best model for these data.

Regarding the choice of p in each model, according to Xiao et al. (2015) one of the most common values used in quantile regression for risk management is 0.95. However, since the goal here is to not only estimate extreme losses in the right tail, but to project a region where future observations are more likely to be, this value will be split and the focus is going to be on the 0.025 and 0.975 quantiles. Also, in order to have a central measure to compare with the mean models, we will also work with the median, $p = 0.50$.

Looking at the estimation of quantiles, it was possible to see that the differences among models became clearer. For all models $1,500,000$ iterations were run with a burn-in of $600,000$ and thinning of 10.

In the log-anova, as simulations have shown (see appendix 5.C), both the last line α and the last column β are very difficult to estimate and their distributions contain a lot of variability. This, in turn, became a problem for the 0.025, and, 0.975 quantiles. In the last line for example, forecasts from the 0.025 were very close to zero, while the 0.975 had some values as large as 1×10^{10}. It is almost like the model was telling that due to uncertainty the reserves should be somewhere between zero and infinity. Figure 5.3 a shows an example of the fitted values for the third line of the triangle using the log-anova, the only value not shown is the one corresponding to the last column for the 0.975 quantile, because it is so large that it would not appear in the



(a) Log-Anova (b) Log-Ancova (c) Log-SS

Figure 5.3: Quantile estimation for each model for line 3 of the triangle, the colored lines represent the mean of the fitted values for quantiles $0.025, 0.5$and, 0.975.

plot ($\$2,841,263.76$).

The log-ancova had almost the same issues as the previous model. The only difference is that in this case there is only one parameter for the lines, which means that all data is used to estimate this single parameter, α. Therefore, the huge uncertainty that was present both in the last line and last column of the log-anova just manifests itself in the last column here. Figure 5.3 b shows an example of the fitted values for the third line of the triangle, again not including the last value of the 0.975 quantile, for the same reason as before ($\$1,954,515.72$).

Finally, in the log-state-space, even though the evolution equations in both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ represent an increase in the variance over time, the simulations have shown (see appendix $5.C$) that the variability in the distributions of the parameters in the last line and last column of the triangle is nowhere near the variability in the previous models. This essentially means that those extremely small and large values present in the log-anova and log-ancova do not exist in this model. Figure 5.3 c shows an example of the fitted values for the third line of the triangle, this time showing the last value of the 0.975 quantile ($\$3,609.61$).

Following those brief explanations, we proceed now to show the reserves generated by each of these models, following the same reasoning as in chapter 4. Figure 5.4 shows histograms of these reserves provided by the log-anova model for quantiles 0.025, 0.50, and 0.975. Also, table 5.2 displays median and 95% interval associated with these reserves. The histograms show the summation of the diagonal marked in red in table 3.2. The lower quantile has a very interesting form and the values seem to be consistent. However, the upper quantile show signs of the variability and uncertainty discussed before, with the 95% interval displaying really large values. For instance, note that the upper limit, for the 0.975 quantile, has a value that is not even shown on the x-axis of the histogram, because the axis had to be cut in order to have a good visualization.

Figure 5.5 shows histograms for next year reserves for the log-ancova and table 5.3 presents the median and 95% credible intervals for the same model. It shows a drastic decrease in the variability when compared to the previous model. Especially in the upper quantile, just by taking a quick look at the histograms it is possible to see that. Also, the median for the 0.975 quantile

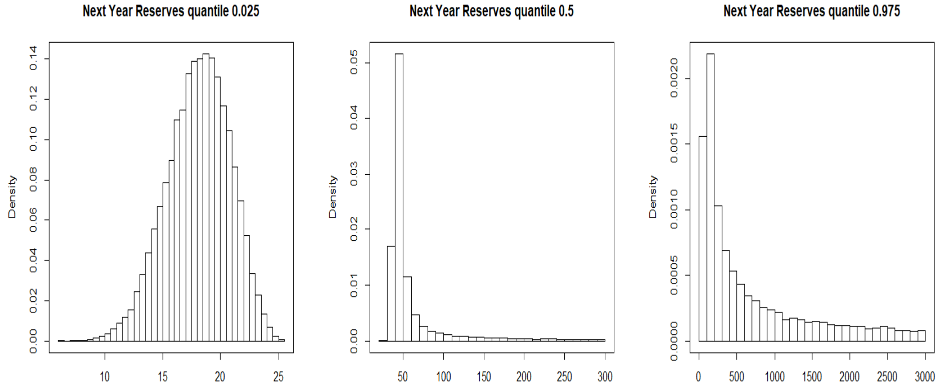


Figure 5.4: Quantile estimation of next year reserves using log-anova, assuming the errors to follow asymmetric Laplace distributions fiting data from 4.1. The histogram in the left is for $p = 0.025$, in the center $p = 0.5$ and in the right $p = 0.975$. Values were divided by $1,000$.

Table 5.2: Median and 95% quantile estimation of next year reserves using log-anova. Assuming errors to follow asymmetric Laplace distribution fitting data from 4.1. The upper and lower bounds indicate the extremes of the credible interval.

	Log-Anova		
Quantiles	Lower Bound	Median	Upper Bound
0.025	$12,497.18$	$18,218.21$	$22,914.86$
0.50	$34,971.42$	$45,610.43$	$210,661.43$
0.975	$76,827.23$	$422,783.54$	$4,447,938.55$

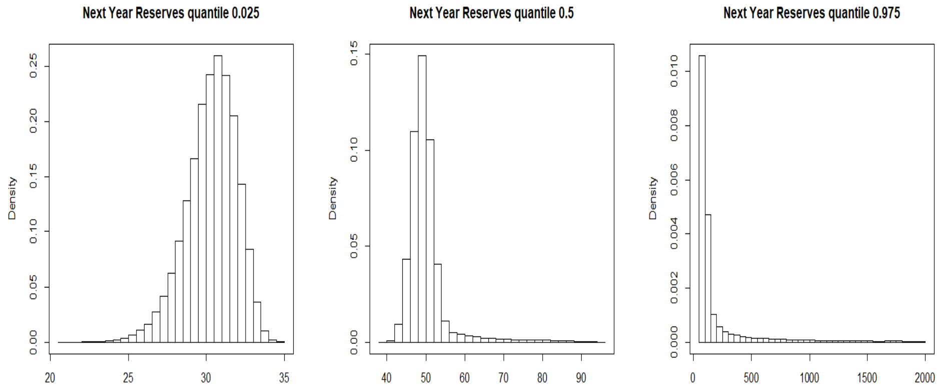


Figure 5.5: Quantile estimation of next year reserves using log-ancova, assuming the errors to follow asymmetric Laplace distributions fiting data from 4.1. The histogram in the left is for $p = 0.025$, in the center $p = 0.5$ and in the right $p = 0.975$. Values were divided by $1,000$.

Table 5.3: Median and 95% quantile estimation of next year reserves using log-ancova. Assuming errors to follow asymmetric Laplace distribution fitting data from 4.1. The upper and lower bounds indicate the extremes of the credible interval.

		Log-Ancova	
Quantiles	Lower Bound	Median	Upper Bound
0.025	$26,678.42$	$30,462.40$	$32,986.70$
0.50	$44,214.70$	$49,156.41$	$67,856.41$
0.975	$88,857.77$	$99,558.63$	$1,898,894.73$

was reduced from over $400,000$ to less than $100,000$. This might be due to the fact that α is better estimated in this model.

Lastly, let us have a look at the log-state-space model. Figure 5.6 shows histograms for next year reserves and table 5.4 presents the median and 95% interval. First of all, just by glancing at the histograms it is clear that the difference is outstanding. The quantiles distributions are way more concentrated and the intervals do not show values extremely large in the upper bound as in the previous models.

Interestingly, even though each model presents certain particularities, the median values

Table 5.4: Median and 95% quantile estimation of next year reserves using log-state-space. Assuming errors to follow asymmetric Laplace distribution fitting data from 4.1. The upper and lower bounds indicate the extremes of the credible interval.

		Log-State-Space	
Quantiles	Lower Bound	Median	Upper Bound
0.025	$8,785.25$	$24,420.35$	$49,415.87$
0.50	$24,045.24$	$44,566.43$	$70,434.13$
0.975	$59,424.89$	$84,088.97$	$111,273.19$

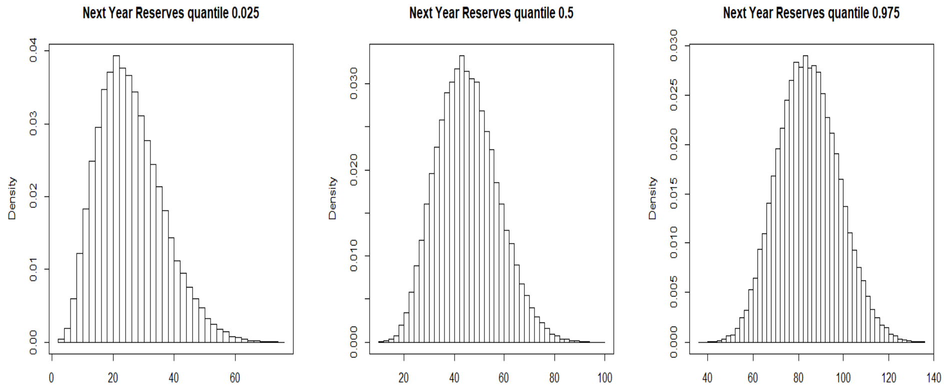


Figure 5.6: Quantile estimation of next year reserves using log-state-space, assuming the errors to follow asymmetric Laplace distributions fiting data from 4.1. The histogram in the left is for $p = 0.025$, in the center $p = 0.5$ and in the right $p = 0.975$. Values were divided by $1,000$.

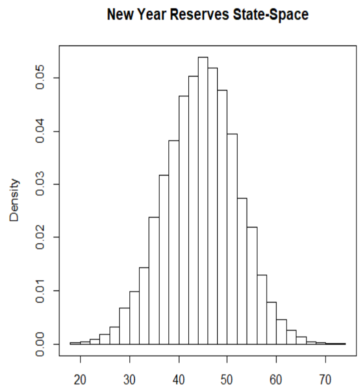


Figure 5.7: Next year reserves estimated using the log-state-space, assuming the errors to follow a student's t distribution fitting a data from 4.1

Table 5.5: Mean estimation and 95% interval of next year reserves using log-state-space assuming errors student's t fitting data from 4.1.

	Log-State-Space		
	Lower Bound	Median	Upper Bound
Mean Model	$29,805.17$	$44,828.75$	$58,987.75$

are not all that different, especially when looking at the 0.50 quantile. This might mean that when using this models to only estimate a central measure for the reserves, as it is done in the conditional mean case, they may provide reasonably close forecasts.

Finally, let us compare results from table 5.4 with the same reserves estimates provided by the log-state-space with student's t errors, since there is evidence that this model is the best for these data, see table 4.3. Comparing it with the histogram of the 0.50 quantile generated from the same model with asymmetric Laplace distribution, it is possible to see that the histogram of the mean model (see figure 4.4) looks more symmetric, while the other one seems to have a bit of a right tail. This is confirmed by the values of the median and 95% credible interval in table 5.5, where the most different value is the upper bound of the interval, while the median is, indeed, almost the same.

5.5 Discussion

This chapter have presented the models with errors following asymmetric Laplace distributions. In the applications section, simulations with artificial data have been carried out along with a real-world application using data from Choy et al. (2016). Showing that the asymmetric Laplace can recover the quantiles of the underlying distribution could represent a huge saving of time during the modeling process, once instead of trying three or four distributions before deciding which one is better, one can simply fit the data with the ALD. This is shown in tables 5.4 and 5.5, where there is only a slight difference between the median of the log-state-space fitted with student's t ($\$44,828.75$) and asymmetric Laplace ($\$44,566.43$) distributions.

Therefore, the results presented here concur with the initial idea that one can use the quantile model to estimate a central measure for future reserves as well as extreme quantiles of the claims distribution. This makes the models described in this work more robust than the traditional ones. Also, the estimation of an extreme upper quantile, such as the 0.975, can give practitioners a better understanding of their insurance portfolio exposure, thus enabling them to transfer those risks in order to reduce the chances of catastrophic losses for the company.

Appendix

5.A Full Conditional Posterior Distributions

5.A.1 Log-Anova

The log-anova has the following formulation,

$$
\begin{gathered}
Z_{ij} = \mu_{ij} + \theta\nu_{ij} + \tau\sqrt{\sigma\nu_{ij}}u_{ij}, \\
\mu_{ij} = \mu + \alpha_i + \beta_j, \\
\sum_{i=1}^{n} \alpha_i = 0 \text{ and } \sum_{j=1}^{n} \beta_j = 0, \\
\nu_{ij} \sim \mathcal{E}(\sigma), \text{ and } u_{ij} \sim \mathcal{N}(0,1), \\
Z_{ij} \sim \mathcal{N}\left(\mu_{ij} + \theta\nu_{ij},\ \tau^2\sigma\nu_{ij}\right),
\end{gathered}
$$

For μ, considering its prior as $\mu \sim \mathcal{N}(0, \sigma_\mu^2)$

$$
\begin{aligned}
p\left(\mu|\boldsymbol{z},\sigma,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},p\right) &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}(z_{ij}-\mu-\alpha_i-\beta_j-\theta\nu_{ij})^2\right\} \times \exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(-2z_{ij}\mu+\mu^2+2\mu\alpha_i+2\mu\beta_j+2\mu\theta\nu_{ij}\right)\right\} \times \exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\mu^2}\right)\mu^2-2\mu\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\alpha_i-\beta_j-\theta\nu_{ij}\right)\right)\right]\right\}
\end{aligned}
$$

Completing squares it is possible to see that,

$$
\mu \mid \boldsymbol{z},\sigma,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},p \sim \mathcal{N}\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\alpha_i-\beta_j-\theta\nu_{ij}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\mu{}^2}}, + \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\mu}{}^2}\right)
$$

For σ, considering its prior as $\sigma \sim \mathcal{IG}\left(a,b\right)$ and $N = \frac{n(n+1)}{2}$

$$
\begin{aligned}
p\left(\sigma|\boldsymbol{z},\mu,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},p\right) &\propto \left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left(\frac{1}{\sigma^2}\right)^{1/2}\exp\left\{-\frac{1}{2\tau^2\sigma\nu_{ij}}(z_{ij}-\mu-\alpha_i-\beta_j-\theta\nu_{ij})^2\right\}\right\} \\
&\times\left(\frac{1}{\sigma}\right)^{a+1}\exp\left\{-\frac{b}{\sigma}\right\}\times\left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\frac{1}{\sigma}\ \exp\left\{-\frac{\nu_{ij}}{\sigma}\right\}\right\} \\
&\propto\left(\frac{1}{\sigma}\right)^{(N+a)+1}\exp\left\{-\frac{1}{\sigma}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{2\tau^2\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_j-\theta\nu_{ij}\right)^2+\nu_{ij}\right)+b\right]\right\}
\end{aligned}
$$

Then, it is clear that

$$
\sigma|\boldsymbol{z},\mu,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},p\sim\mathcal{IG}\left(\frac{N}{2}+a,\ \left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{2\tau^2\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_j-\theta\nu_{ij}\right)^2+\nu_{ij}\right)+b\right)
$$

For a particular α_i, considering its prior as $\alpha_i\sim\mathcal{N}(0,\sigma^2_\alpha)$

$$
\begin{aligned}
p\left(\alpha_i|\boldsymbol{z},\mu,\sigma,\boldsymbol{\beta},\boldsymbol{\nu},p\right) &\propto \exp\left\{-\frac{1}{2}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_j-\theta\nu_{ij}\right)^2\right\}\times\exp\left\{-\frac{{\alpha_i}^2}{2\sigma^2_\alpha}\right\} \\
&\propto\exp\left\{-\frac{1}{2}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(-2z_{ij}\alpha_i+{\alpha_i}^2+2\alpha_i\mu+2\alpha_i\beta_j-2\alpha_i\theta\nu_{ij}\right)\right\}\times\exp\left\{-\frac{{\alpha_i}^2}{2\sigma^2_\alpha}\right\} \\
&\propto\exp\left\{-\frac{1}{2}\left[\left(\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma^2_\alpha}\right){\alpha_i}^2-2\alpha_i\left(\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\beta_j\right)\right)\right]\right\}
\end{aligned}
$$

Once again, completing squares

$$
\alpha_i|\boldsymbol{z},\mu,\sigma,\boldsymbol{\beta},\boldsymbol{\nu},p\sim\ \mathcal{N}\left(\frac{\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\beta_j-\theta\nu_{ij}\right)}{\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{{\sigma_\alpha}^2}},\ \frac{1}{\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{{\sigma_\alpha}^2}}\right),
$$

for $i=1,\ldots,n-1$.

For a particular β_j, considering its prior as $\beta_j\sim\mathcal{N}(0,\sigma^2_\beta)$

$$
\begin{aligned}
p\left(\beta_j|\boldsymbol{z},\mu,\sigma,\boldsymbol{\alpha},\boldsymbol{\nu},p\right) &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n-j+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_j-\theta\nu_{ij}\right)^2\right\}\times\exp\left\{-\frac{{\beta_j}^2}{2\sigma^2_\beta}\right\} \\
&\propto\exp\left\{-\frac{1}{2}\sum_{i=1}^{n-j+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(-2z_{ij}\beta_j+{\beta_j}^2+2\beta_j\mu+2\beta_j\alpha_i+2\beta_j\theta\nu_{ij}\right)\right\}\times\exp\left\{-\frac{\beta_j 2}{2\sigma^2_\beta}\right\} \\
&\propto\exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n-j+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma^2_\beta}\right){\beta_j}^2-2\beta_j\left(\sum_{i=1}^{n-j+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\theta\nu_{ij}\right)\right)\right]\right\}
\end{aligned}
$$

Completing squares

$$\beta_j|\boldsymbol{z},\mu,\sigma,\boldsymbol{\alpha},\boldsymbol{\nu},p\sim\ \mathcal{N}\left(\frac{\sum_{i=1}^{n-j+1}\ \frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\theta\nu_{ij}\right)}{\sum_{i=1}^{n-j+1}\ \frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\beta{}^2}},\ \frac{1}{\sum_{i=1}^{n-j+1}\ \frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\beta{}^2}}\right),$$

for $j=1,\ldots,n-1$.

For ν_{ij}, considering its prior as $\nu_{ij}\sim\mathcal{E}(\sigma)$

$$\begin{aligned}p\left(\nu_{ij}|z_{ij},\mu,\sigma,\alpha_i,\beta_j,p\right)&\propto\left(\frac{1}{\nu_{ij}}\right)^{1/2}\exp\left\{-\frac{1}{2\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_j-\theta\nu_{ij}\right)^2\right\}\times\frac{1}{\sigma}\exp\left\{-\frac{\nu_{ij}}{\sigma}\right\}\\&\propto\left(\frac{1}{\nu_{ij}}\right)^{1/2}\exp\left\{-\frac{1}{2}\left[\frac{\left(z_{ij}-\mu-\alpha_i-\beta_j\right)^2}{\tau^2\sigma\nu_{ij}}+\left(\frac{\theta}{\tau^2\sigma}+\frac{2}{\sigma}\right)\nu_{ij}\right]\right\}\end{aligned}$$

Then, it is easy to see that

$$\nu_{ij}|z_{ij},\mu,\sigma,\alpha_i,\beta_j,p\sim\mathcal{GIG}\left(\frac{1}{2},\ \frac{\left(z_{ij}-\mu-\alpha_i-\beta_j\right)^2}{\tau^2\sigma},\left(\frac{\theta}{\tau^2\sigma}+\frac{2}{\sigma}\right)\right)$$

for$i=1,\ldots,n$ and $j=1,\ldots,n-i+1$.

5.A.2 Log-Ancova

The log-ancova has the following formulation,

$$\begin{gathered}Z_{ij}=\mu_{ij}+\theta\nu_{ij}+\tau\sqrt{\sigma\nu_{ij}}u_{ij},\\\mu_{ij}=\mu+i\alpha+\beta_j,\\\sum_{j=1}^{n}\beta_j=0,\\\nu_{ij}\sim\mathcal{E}(\sigma),\ \text{and}\ u_{ij}\sim\mathcal{N}(0,1),\\Z_{ij}\sim\mathcal{N}\left(\mu_{ij}+\theta\nu_{ij},\ \tau^2\sigma\nu_{ij}\right),\end{gathered}$$

Therefore, for μ, considering its prior as $\mu\sim\mathcal{N}(0,\sigma_\mu^2)$

$$\begin{aligned}p\left(\mu|\boldsymbol{z},\sigma,\alpha,\boldsymbol{\beta},\boldsymbol{\nu},p\right)&\propto\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-i\alpha-\beta_j-\theta\nu_{ij}\right)^2\right\}\times\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\}\\&\propto\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\frac{\left(-2z_{ij}\mu+\mu^2+2\mu\ i\alpha+2\mu\beta_j+2\mu\theta\nu_{ij}\right)^2}{2}\right\}\times\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\}\\&\propto\exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\mu^2}\right)\mu^2-2\mu\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-i\alpha-\beta_j-\theta\nu_{ij}\right)\right)\right]\right\}\end{aligned}$$

Completing squares it is possible to see that,

$$\mu \mid \boldsymbol{z}, \sigma, \alpha, \boldsymbol{\beta}, \boldsymbol{\nu}, p \sim \mathcal{N}\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(Z_{ij}-i\alpha-\beta_j-\theta\nu_{ij}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\mu{}^2}}, +\frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\mu}^2}\right)$$

For σ, considering its prior as $\sigma \sim \mathcal{IG}\left(a, b\right)$ and $N = \frac{n(n+1)}{2}$

$$\begin{aligned}
p\left(\sigma|\boldsymbol{z}, \mu, \alpha, \boldsymbol{\beta}, \boldsymbol{\nu}, p\right) &\propto \left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left(\frac{1}{\sigma}\right)^{1/2}\exp\left\{-\frac{1}{2\tau^2\sigma\nu_{ij}}(z_{ij}-\mu-i\alpha-\beta_j-\theta\nu_{ij})^2\right\}\right\} \\
&\times \left(\frac{1}{\sigma}\right)^{a+1}\exp\left\{-\frac{b}{\sigma}\right\} \times \left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\frac{1}{\sigma}\exp\left\{-\frac{\nu_{ij}}{\sigma}\right\}\right\} \\
&\propto \left(\frac{1}{\sigma}\right)^{(N/2\;+\;a)+1}\exp\left\{-\frac{1}{\sigma}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{2\tau^2\nu_{ij}}\left(z_{ij}-\mu-i\alpha-\beta_j-\theta\nu_{ij}\right)^2+\nu_{ij}\right)+b\right]\right\}
\end{aligned}$$

Then, it is clear that

$$\sigma|\boldsymbol{z}, \mu, \alpha, \boldsymbol{\beta}, \boldsymbol{\nu}, p \sim \mathcal{IG}\left(\frac{N}{2}+a,\ \left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{2\tau^2\nu_{ij}}\left(z_{ij}-\mu-i\alpha-\beta_j-\theta\nu_{ij}\right)^2+\nu_{ij}\right)+b\right)$$

For α, considering its prior as $\alpha \sim \mathcal{N}(0, \sigma_\alpha^2)$

$$\begin{aligned}
p\left(\alpha|\boldsymbol{z}, \mu, \sigma, \boldsymbol{\beta}, \boldsymbol{\nu}, p\right) &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-i\alpha-\beta_j-\theta\nu_{ij}\right)^2\right\} \times \exp\left\{-\frac{\alpha^2}{2\sigma_\alpha^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(-2z_{ij}\ i\alpha+(i\alpha)^2+2i\alpha\mu+2i\alpha\beta_j+2i\alpha\theta\nu_{ij}\right)\right\} \\
&\times \exp\left\{-\frac{\alpha 2}{2\sigma_\alpha^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{i^2}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\alpha^2}\right)\alpha^2-2\alpha\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{i}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\beta_j-\theta\nu_{ij}\right)\right)\right]\right\}
\end{aligned}$$

Once again, completing squares

$$\alpha|\ \boldsymbol{z}, \mu, \sigma, \boldsymbol{\beta}, \boldsymbol{\nu}, p \sim \mathcal{N}\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{i}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\beta_j-\theta\nu_{ij}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{i^2}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\alpha{}^2}}, \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{i^2}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\alpha{}^2}}\right).$$

For a particular β_j, considering its prior as $\beta_j \sim \mathcal{N}(0, \sigma_\beta^2)$

$$
\begin{aligned}
p\left(\beta_j|\boldsymbol{z},\mu,\sigma,\alpha,\boldsymbol{\nu},p\right) &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n-j+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-i\alpha-\beta_j-\theta\nu_{ij}\right)^2\right\}\times\exp\left\{-\frac{\beta_j{}^2}{2\sigma_\beta^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n-j+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(-2z_{ij}\beta_j+\beta_j{}^2+2\beta_j\mu+2\beta_j\ i\alpha+2\beta_j\theta\nu_{ij}\right)\right\} \\
&\times\exp\left\{-\frac{\beta_j 2}{2\sigma_\beta^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n-j+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\beta^2}\right)\beta_j{}^2-2\beta_j\left(\sum_{i=1}^{n-j+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-i\alpha-\theta\nu_{ij}\right)\right)\right]\right\}
\end{aligned}
$$

Completing squares

$$
\beta_j|\boldsymbol{z},\mu,\sigma,\alpha,\boldsymbol{\nu},p\sim\ \mathcal{N}\left(\frac{\sum_{i=1}^{n-j+1}\ \frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-i\alpha-\theta\nu_{ij}\right)}{\frac{1}{\sigma^2}\sum_{i=1}^{n-j+1}\ \frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\beta{}^2}},\ \frac{1}{\sum_{i=1}^{n-j+1}\ \frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\beta{}^2}}\right),
$$

for $j=1,\ldots,n-1$.

For ν_{ij}, considering its prior as $\nu_{ij}\sim\mathcal{E}(\sigma)$

$$
\begin{aligned}
p\left(\nu_{ij}|z_{ij},\mu,\sigma,\alpha,\beta_j,p\right) &\propto \left(\frac{1}{\nu_{ij}}\right)^{1/2}\exp\left\{-\frac{1}{2\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-i\alpha-\beta_j-\theta\nu_{ij}\right)^2\right\}\times\frac{1}{\sigma}\exp\left\{-\frac{\nu_{ij}}{\sigma}\right\} \\
&\propto \left(\frac{1}{\nu_{ij}}\right)^{1/2}\exp\left\{-\frac{1}{2}\left[\frac{\left(z_{ij}-\mu-i\alpha-\beta_j\right)^2}{\tau^2\sigma\nu_{ij}}+\left(\frac{\theta}{\tau^2\sigma}+\frac{2}{\sigma}\right)\nu_{ij}\right]\right\}
\end{aligned}
$$

Then, it is easy to see that

$$
\nu_{ij}|z_{ij},\mu,\sigma,\alpha,\beta_j,p\sim\mathcal{GIG}\left(\frac{1}{2},\ \frac{\left(z_{ij}-\mu-i\alpha-\beta_j\right)^2}{\tau^2\sigma},\left(\frac{\theta}{\tau^2\sigma}+\frac{2}{\sigma}\right)\right)
$$

for$i=1,\ldots,n$ and $j=1,\ldots,n-i+1$.

5.A.3 Log-State-Space

As already describe the log-state-space has the following formulation,

$$
\begin{gathered}
Z_{ij}=\mu_{ij}+\theta\nu_{ij}+\tau\sqrt{\sigma\nu_{ij}}u_{ij}, \\
\mu_{ij}=\mu+\alpha_i+\beta_{ij}, \\
\alpha_1=0,\ \text{and}\ \boldsymbol{\beta_{i1}}=\mathbf{0}, \\
\nu_{ij}\sim\mathcal{E}(\sigma),\ \text{and}\ u_{ij}\sim\mathcal{N}(0,1), \\
Z_{ij}\sim\mathcal{N}\left(\mu_{ij}+\theta\nu_{ij},\ \tau^2\sigma\nu_{ij}\right),
\end{gathered}
$$

where, $\mathcal{E}(.)$ is the exponential distribution and $\mathcal{N}(.,.)$ in the normal distribution. Therefore, for μ, considering its prior as $\mu\sim\mathcal{N}(0,\sigma_\mu^2)$

$$
\begin{aligned}
p\left(\mu|\boldsymbol{z},\sigma,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},p\right) &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_{ij}-\theta\nu_{ij}\right)^2\right\}\times\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(-2z_{ij}\mu+\mu^2+2\mu\alpha_i+2\mu\beta_{ij}+2\mu\theta\nu_{ij}\right)\right\}\times\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\mu^2}\right)\mu^2-2\mu\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\alpha_i-\beta_{ij}-\theta\nu_{ij}\right)\right)\right]\right\}
\end{aligned}
$$

Completing squares it is possible to see that,

$$
\mu|\boldsymbol{z},\sigma,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},p\sim\ \mathcal{N}\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(Z_{ij}-\alpha_i-\beta_{ij}-\theta\nu_{ij}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\mu{}^2}},+\frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\mu}^2}\right)
$$

For σ, considering its prior as $\sigma\sim\mathcal{IG}\left(a,b\right)$ and $N=\frac{n(n+1)}{2}$

$$
\begin{aligned}
p\left(\sigma|\boldsymbol{z},\mu,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},p\right) &\propto \left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left(\frac{1}{\sigma}\right)^{1/2}\exp\left\{-\frac{1}{2\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_{ij}-\theta\nu_{ij}\right)^2\right\}\right\}\\
&\times\left(\frac{1}{\sigma}\right)^{a+1}\exp\left\{-\frac{b}{\sigma}\right\}\times\left\{\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\frac{1}{\sigma}\ \exp\left\{-\frac{\nu_{ij}}{\sigma}\right\}\right\}\\
&\propto\left(\frac{1}{\sigma}\right)^{(N+a)+1}\exp\left\{-\frac{1}{\sigma}\left(\left[\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{2\tau^2\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_{ij}-\theta\nu_{ij}\right)^2+\nu_{ij}\right]+b\right)\right\}
\end{aligned}
$$

Then, it is clear that

$$
\sigma|\boldsymbol{z},\mu,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},p\sim\mathcal{IG}\left(N+a,\ \left(\left[\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{1}{2\tau^2\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_{ij}-\theta\nu_{ij}\right)^2+\nu_{ij}\right]+b\right)\right)
$$

For a particular α_i, considering its prior as $\alpha_i\sim\mathcal{N}(\alpha_{i-1},\sigma_\alpha^2)$

$$
\begin{aligned}
p\left(\alpha_i|\boldsymbol{z},\mu,\sigma,\boldsymbol{\beta},\boldsymbol{\nu},p\right) \propto & \exp\left\{-\frac{1}{2}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_{ij}-\theta\nu_{ij}\right)^2\right\} \\
& \times\exp\left\{-\frac{\left(\alpha_i-\alpha_{i-1}\right)^2}{2\sigma_\alpha^2}\right\}\times\exp\left\{-\frac{\left(\alpha_{i+1}-\alpha_i\right)^2}{\sigma_\alpha^2}\right\} \\
& \propto\exp\left\{-\frac{1}{2}\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(-2z_{ij}\alpha_i+{\alpha_i}^2+2\alpha_i\mu+2\alpha_i\beta_{ij}+2\alpha_i\theta\nu_{ij}\right)\right\} \\
& \times\exp\left\{-\frac{\left({\alpha_i}^2-2\alpha_{i-1}\alpha_i\right)}{2\sigma_\alpha^2}\right\}\times\exp\left\{-\frac{\left(-2\alpha_{i+1}\alpha_i+{\alpha_i}^2\right)}{2\sigma_\alpha^2}\right\} \\
& \propto\exp\left\{-\frac{1}{2}\left[\left(\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{2}{\sigma_\alpha^2}\right){\alpha_i}^2-\right.\right. \\
& \left.\left. 2\alpha_i\left(\frac{\alpha_{i-1}+\alpha_{i+1}}{\sigma_\alpha^2}+\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\beta_{ij}-\theta\nu_{ij}\right)\right)\right]\right\}
\end{aligned}
$$

Once again, completing squares

$$
\alpha_i|\boldsymbol{z},\mu,\sigma,\boldsymbol{\beta},\boldsymbol{\nu},p\sim\mathcal{N}\left(\frac{\frac{\alpha_{i-1}+\alpha_{i+1}}{\sigma_\alpha^2}+\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\beta_{ij}-\theta\nu_{ij}\right)}{\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{2}{\sigma_\alpha^2}},\frac{1}{\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{2}{\sigma_\alpha^2}}\right),
$$

for $i=2,\ldots,n-1$.

It is also straightforward to see that for $i=n$

$$
\alpha_i|\boldsymbol{z},\mu,\sigma,\boldsymbol{\beta},\boldsymbol{\nu},p\sim\mathcal{N}\left(\frac{\frac{\alpha_{i-1}}{\sigma_\alpha^2}+\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\beta_{ij}-\theta\nu_{ij}\right)}{\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\alpha^2}},\frac{1}{\sum_{j=1}^{n-i+1}\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\alpha^2}}\right),
$$

for $i=n$.

For a particular β_{ij}, considering its prior as $\beta_{ij}\sim\mathcal{N}(\beta_{i-1j},\sigma_\beta^2)$

$$
\begin{aligned}
p\left(\beta_{ij}|z_{ij},\mu,\sigma,\alpha_i,\nu_{ij},p\right)\propto & \exp\left\{-\frac{1}{2\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_{ij}-\theta\nu_{ij}\right)^2\right\} \\
& \times\exp\left\{-\frac{\left(\beta_{ij}-\beta_{i-1j}\right)^2}{2\sigma_\beta^2}\right\}\times\exp\left\{-\frac{\left(\beta_{i+1j}-\beta_{ij}\right)^2}{2\sigma_\beta^2}\right\} \\
& \propto\exp\left\{-\frac{1}{2}\frac{1}{\tau^2\sigma\nu_{ij}}\left(-2z_{ij}\beta_{ij}+{\beta_{ij}}^2+2\beta_{ij}\mu+2\beta_{ij}\alpha_i+2\beta_{ij}\theta\nu_{ij}\right)\right\} \\
& \times\exp\left\{-\frac{\left({\beta_{ij}}^2-2\beta_{ij}\beta_{i-j}\right)}{2\sigma_\beta^2}\right\}\times\exp\left\{-\frac{\left(-2\beta_{i+1j}\beta_{ij}+{\beta_{ij}}^2\right)}{2\sigma_\beta^2}\right\} \\
& \propto\exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{2}{\sigma_\beta^2}\right){\beta_{ij}}^2-\right.\right. \\
& \left.\left. 2\beta_{ij}\left(\frac{\left(\beta_{i+1j}+\beta_{i-1j}\right)}{\sigma_\beta^2}+\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\theta\nu_{ij}\right)\right)\right]\right\}
\end{aligned}
$$

Completing squares

$$\beta_{ij}|z_{ij},\mu,\sigma,\alpha_i,\nu_{ij},p \sim \mathcal{N}\left(\frac{\frac{\beta_{i-1j}+\beta_{i+1j}}{\sigma_\beta^2}+\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\theta\nu_{ij}\right)}{\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{2}{\sigma_\beta^2}}, \frac{1}{\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{2}{\sigma_\beta^2}}\right),$$

for $i=1,\ldots,n-j$ and $j=2,\ldots,n$.

Again, it is straightforward to see that

$$\beta_{ij}|z_{ij},\mu,\sigma,\alpha_i,\nu_{ij},p \sim \mathcal{N}\left(\frac{\frac{\beta_{i-1j}}{\sigma_\beta^2}+\frac{1}{\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\theta\nu_{ij}\right)}{\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\beta^2}}, \frac{1}{\frac{1}{\tau^2\sigma\nu_{ij}}+\frac{1}{\sigma_\beta^2}}\right),$$

for $i=n-j+1$ and $j=2,\ldots,n$.

For ν_{ij}, considering its prior as $\nu_{ij}\sim\mathcal{E}(\sigma)$

$$\begin{aligned} p\left(\nu_{ij}|z_{ij},\mu,\sigma,\alpha_i,\beta_{ij},p\right) &\propto \left(\frac{1}{\nu_{ij}}\right)^{1/2}\exp\left\{-\frac{1}{2\tau^2\sigma\nu_{ij}}\left(z_{ij}-\mu-\alpha_i-\beta_{ij}-\theta\nu_{ij}\right)^2\right\}\times\frac{1}{\sigma}\exp\left\{-\frac{\nu_{ij}}{\sigma}\right\} \\ &\propto \left(\frac{1}{\nu_{ij}}\right)^{1/2}\exp\left\{-\frac{1}{2}\left[\frac{\left(z_{ij}-\mu-\alpha_i-\beta_{ij}\right)^2}{\tau^2\sigma\nu_{ij}}+\left(\frac{\theta}{\tau^2\sigma}+\frac{2}{\sigma}\right)\nu_{ij}\right]\right\} \end{aligned}$$

Then, it is easy to see that

$$\nu_{ij}|z_{ij},\mu,\sigma,\alpha_i,\beta_{ij},p \sim \mathcal{GIG}\left(\frac{1}{2},\ \frac{\left(z_{ij}-\mu-\alpha_i-\beta_{ij}\right)^2}{\tau^2\sigma},\left(\frac{\theta}{\tau^2\sigma}+\frac{2}{\sigma}\right)\right)$$

for$i=1,\ldots,n$ and $j=1,\ldots,n-i+1$.

The variance of $\boldsymbol{\beta}$ will receive a uniform prior $\sigma_\beta^2\sim\mathcal{U}(0,A)$, with $A\to\infty$

$$\begin{aligned} p\left(\sigma_\beta^2|\boldsymbol{\beta}\right) &\propto \prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left(\frac{1}{\sigma_\beta^2}\right)^{1/2}\exp\left\{-\frac{\left(\beta_{ij}-\beta_{i-1j}\right)^2}{2\sigma_\beta^2}\right\} \\ &\propto \left(\frac{1}{\sigma_\beta^2}\right)^{N/2}\exp\left\{-\frac{1}{\sigma_\beta^2}\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\left(\beta_{ij}-\beta_{i-1j}\right)^2}{2}\right\} \end{aligned}$$

Therefore,

$$\sigma_\beta^2|\boldsymbol{\beta}\sim\mathcal{IG}\left(\frac{N}{2}+1,\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}\frac{\left(\beta_{ij}-\beta_{i-1j}\right)^2}{2}\right)$$

5.B Generalized Inverse Gaussian Distribution

The generalized inverse Gaussian (GIG) distribution has become quite popular for modeling stock prices in financial mathematics (Hörmann and Leydold, 2014). In our case, this distribution shows up in the derivation of some full conditional posterior distributions when we are dealing with errors following an asymmetric Laplace distribution.

Let X be a random variable, such that $X \sim \mathcal{GIG}(\lambda, \psi, \chi)$, then the density function of X is given by

$$f(x|\lambda, \psi, \chi) = \frac{(\chi/\psi)^2}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} \exp\left\{ -\frac{1}{2}(\chi x^{-1} + \psi x) \right\},$$

where $x \in \mathbb{R}$. Parameters λ, ψ and χ, have to satisfy the condition

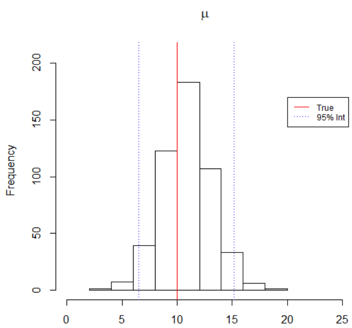
$$\lambda \in \mathbb{R}, \qquad (\psi, \chi) \in \begin{cases} \{(\psi, \chi) : \psi > 0, \chi \geq 0\}, \text{ if } \lambda > 0, \\ \{(\psi, \chi) : \psi > 0, \chi > 0\}, \text{ if } \lambda = 0, \\ \{(\psi, \chi) : \psi \geq 0, \chi > 0\}, \text{ if } \lambda < 0. \end{cases}$$

Also, K_λ is a modified Bessel function of the third kind. In the special case where $\lambda = \frac{1}{2}$, as it happens to be all the cases we have worked with, this function has the form

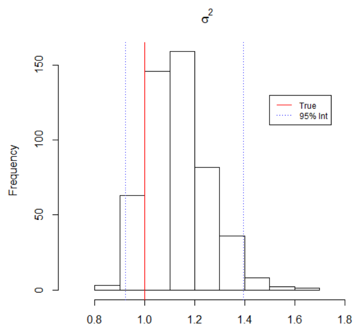
$$K_{\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2}} e^{-x} x^{-\frac{1}{2}}, \ x \in \mathbb{R}^+.$$

5.C Simulations - Asymmetric Laplace Models

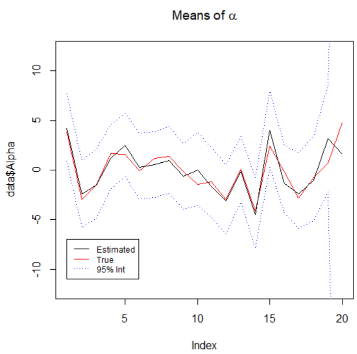
In this appendix, histograms represent the samples of the posterior distributions. In the plots with lines, the ones called Estimated are formed by the means of the posterior distribution of each parameter and the blue dotted lines are credible intervals for each posterior distribution.



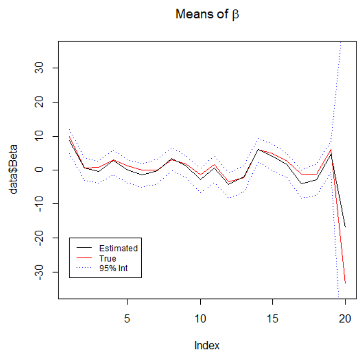
(1) Parameter μ Log-Anova ALD, $p = 0.5$



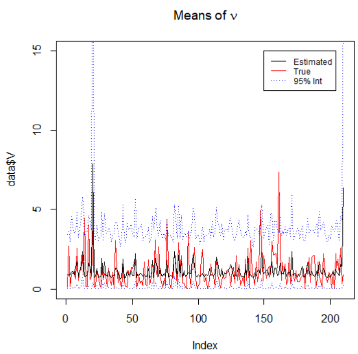
(2) Parameter σ Log-Anova ALD, $p = 0.5$



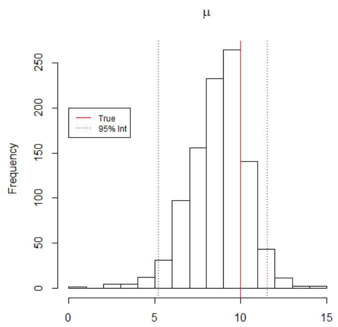
(3) Parameter α Log-Anova ALD, $p = 0.5$



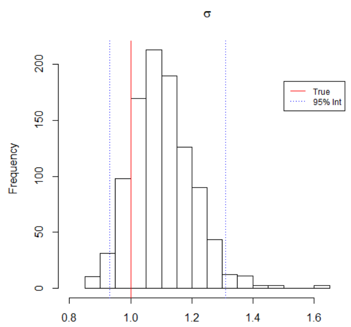
(4) Parameter β Log-Anova ALD, $p = 0.5$



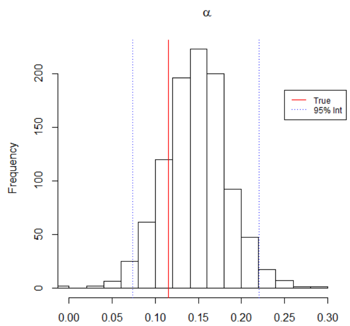
(e) Parameter ν Log-Anova ALD, $p = 0.5$



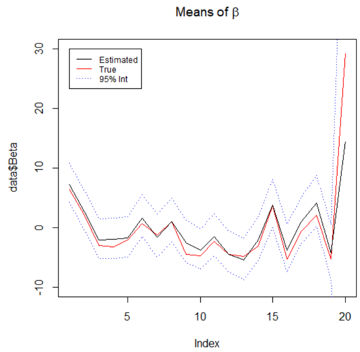
(6) Parameter μ Log-Ancova ALD, $p = 0.5$



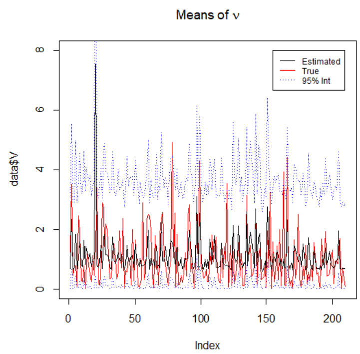
(7) Parameter σ Log-Ancova ALD, $p = 0.5$



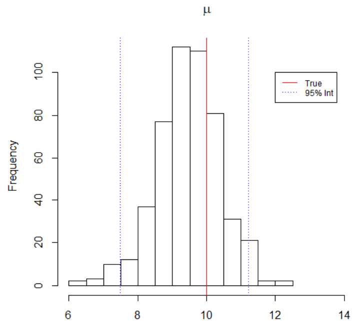
(8) Parameter α Log-Ancova ALD, $p = 0.5$



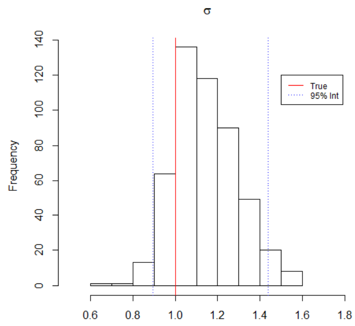
(9) Parameter β Log-Ancova ALD, $p = 0.5$



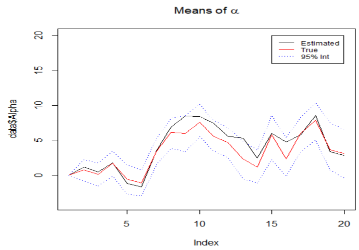
(j) Parameter ν Log-Ancova ALD, $p = 0.5$



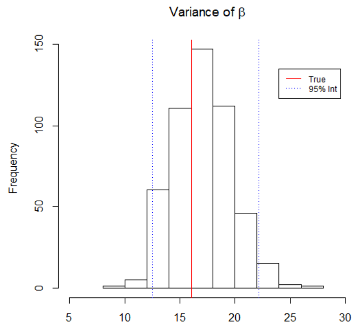
(11) Parameter μ Log-State-Space ALD $p = 0.5$



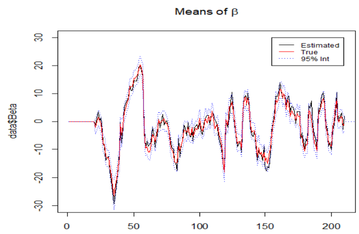
(12) Parameter σ Log-State-Space ALD $p = 0.5$



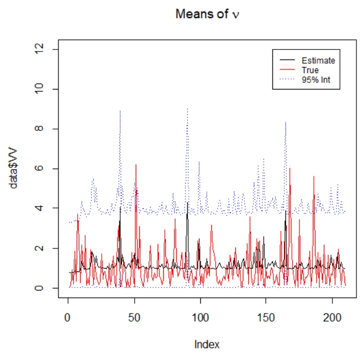
(13) Parameter α Log-State-Space ALD $p = 0.5$



(14) Parameter σ_β^2 Log-State-Space ALD $p = 0.5$



(o) Parameter β Log-State-Space ALD $p = 0.5$



(p) Parameter ν Log-State-Space ALD, $p = 0.5$

Chapter 6

Conclusion

This work had the objective of proposing loss reserving models that are more robust and flexible in the treatment of irregular and extreme claims. In contrast to existing models, which are focused on predicting conditional means as loss reserving estimates, the models proposed are able to deal with irregular claim payment patterns through the estimation of extreme conditional quantiles. These models show some promising characteristics, especially regarding the understanding of risk exposure in insurance portfolios.

More generally, this dissertation extended the work of Choy et al. (2016), using the log analysis of variance, the log analysis of covariance and the log state-space, proposing the use of the asymmetric Laplace distribution (ALD), which is flexible enough to estimate the quantiles of an underlying distribution, regardless of the process that generated it.

Moreover, a Bayesian paradigm using Markov chain Monte Carlo algorithms was implemented with the objective of compensating for uncertainty in the parameter estimation process, since in the loss reserving context there are some parameters that are estimated based on a single observation. Also, using this approach one is able to resort to its computational flexibility to easily produce predictive distributions.

Another important point is that in the applications shown in this work, it was clear that dynamic linear models better capture the variability in the data, when compared to fixed-effect models, which suffer too much with their use of a single parameter for the lag-year effect.

A possible future extension for this work is to propose the use of the number of claims as a co-variable in the model (Ntzoufras and Dellaportas, 2002). In this case the mean function of the log state-space would be given by

$$\mu_{ij} = \mu + \alpha_i + \beta_{ij} + log(\eta_{ij}), \tag{6.1}$$

where η_{ij} would be the number of claims originated in year i and paid with delay $j-1$ years. They would be modeled as $Poisson(\lambda_{ij})$, and the log of λ_{ij} could receive a Poisson regression framework using portfolio specific co-variables.

Bibliography

Andrews, D. F., and Mallows, C. L. (1974), "Scale mixtures of normal distributions," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 99–102.

Bernardi, M., Casarin, R., Maillet, B., and Petrella, L. (2016), "Dynamic Model Averaging for Bayesian Quantile Regression," *arXiv preprint arXiv:1602.00856*, .

Charpentier, A. (2014), *Computational actuarial science with R* CRC Press.

Choy, S. B., and Chan, C. (2003), "Scale mixtures distributions in insurance applications," *ASTIN Bulletin: The Journal of the IAA*, 33(1), 93–104.

Choy, S. B., Chan, J. S., and Makov, U. E. (2016), "Robust Bayesian analysis of loss reserving data using scale mixtures distributions," *Journal of Applied Statistics*, 43(3), 396–411.

De Jong, P., and Zehnwirth, B. (1983), "Claims reserving, state-space models and the Kalman filter," *Journal of the Institute of Actuaries*, 110(1), 157–181.

Fonseca, T. C., Ferreira, M. A., and Migon, H. S. (2008), "Objective Bayesian analysis for the Student-t regression model," *Biometrika*, 95(2), 325–333.

Frees, E. W. (2009), *Regression modeling with actuarial and financial applications* Cambridge University Press.

Gamerman, D., and Lopes, H. F. (2006), *Markov chain Monte Carlo: stochastic simulation for Bayesian inference* Chapman and Hall/CRC.

Gelman, A., Hwang, J., and Vehtari, A. (2014), "Understanding predictive information criteria for Bayesian models," *Statistics and computing*, 24(6), 997–1016.

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

Gelman, A. et al. (2006), "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)," *Bayesian analysis*, 1(3), 515–534.

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

Hörmann, W., and Leydold, J. (2014), "Generating generalized inverse Gaussian random variates," *Statistics and Computing*, 24(4), 547–557.

Kotz, S., Kozubowski, T., and Podgorski, K. (2012), *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance* Springer Science & Business Media.

Kozumi, H., and Kobayashi, G. (2011), "Gibbs sampling methods for Bayesian quantile regression," *Journal of statistical computation and simulation*, 81(11), 1565–1578.

Kremer, E. (1982), "IBNR-claims and the two-way model of ANOVA," *Scandinavian Actuarial Journal*, 1982(1), 47–55.

Migon, H. S., Gamerman, D., and Louzada, F. (2014), *Statistical inference: an integrated approach* CRC press.

Montgomery, D. C. (2017), *Design and analysis of experiments* John wiley & sons.

Ntzoufras, I., and Dellaportas, P. (2002), "Bayesian modelling of outstanding liabilities incorporating claim count uncertainty," *North American Actuarial Journal*, 6(1), 113–125.

Pitselis, G., Grigoriadou, V., and Badounas, I. (2015), "Robust loss reserving in a log-linear model," *Insurance: Mathematics and Economics*, 64, 14–27.

Qin, Z. S., Damien, P., and Walker, S. (2003), Scale mixture models with applications to Bayesian inference,, in *AIP Conference Proceedings*, Vol. 690, AIP, pp. 394–395.

Renshaw, A., and Verrall, R. (1998), "A Stochastic Model Underlying the Chain-Ladder Technique," *British Actuarial Journal*, 4(4), 903–923.

Rizzo, M. L. (2007), *Statistical computing with R* Chapman and Hall/CRC.

Schmidt, K. D., and Seminar, C. L. R. (2006), *Methods and models of loss reserving based on run-off triangles: a unifying survey* Techn. Univ., Inst. für Mathematische Stochastik.

Shaby, B., and Wells, M. T. (2010), "Exploring an adaptive Metropolis algorithm," *Currently under review*, 1, 1–17.

Tsionas, E. G. (2003), "Bayesian quantile inference," *Journal of statistical computation and simulation*, 73(9), 659–674.

Verrall, R. (1996), "Claims reserving and generalised additive models," *Insurance: Mathematics and Economics*, 19(1), 31–43.

Verrall, R. J. (1989), "A state space representation of the chain ladder linear model," *Journal of the Institute of Actuaries*, 116(3), 589–609.

Verrall, R. J. (1991), "Chain ladder and maximum likelihood," *Journal of the Institute of Actuaries*, 118(3), 489–499.

Verrall, R. J. (1994), "A method for modelling varying run-off evolutions in claims reserving," *ASTIN Bulletin: The Journal of the IAA*, 24(2), 325–332.

Vilela, J. M. (2013), *Modelos para Provisão IBNR: Uma Abordagem Bayesiana* Federal University of Rio de Janeiro.

West, M., and Harrison, J. (2006), *Bayesian forecasting and dynamic models* Springer Science & Business Media.

Xiao, Z., Guo, H., and Lam, M. S. (2015), "Quantile regression and Value-at-Risk," , .

Yu, K., and Moyeed, R. A. (2001), "Bayesian quantile regression," *Statistics & Probability Letters*, 54(4), 437–447.

Yu, K., and Zhang, J. (2005), "A three-parameter asymmetric Laplace distribution and its extension," *Communications in Statistics—Theory and Methods*, 34(9-10), 1867–1879.