Bayesian Analysis of Reserving Models and Applications

Xiaodan(Alice) Dong

A thesis submitted in fulfillment of
the requirements for the degree of
Doctor of Philosophy

School of Mathematics and Statistics
University of Sydney

April 2014

Abstract

This thesis focuses on developing models for loss reserving in insurance applications. In the first chapter, a Bayesian approach is presented in order to model heavy tail loss reserving data using the generalized beta distribution of the second kind (GB2) with dynamic mean functions and mixture model representation. The proposed GB2 distribution provides a flexible probability density function, which nests various distributions with light and heavy tails, to facilitate accurate loss reserving in insurance applications. Extending the mean functions to include the state space and threshold models provides a dynamic approach to allow for irregular claims behaviors and legislative change which may occur during the claims settlement period. The mixture of GB2 distributions is proposed as a mean of modeling the unobserved heterogeneity which arises from the incidence of very large claims in the loss reserving data. It is shown through both simulation study and forecasting that model parameters are estimated in high accuracy.

Apart from predicting the expected loss in the future, risk margin estimation is another important aspect of loss reserving. We propose to develop quantile regression to derive risk margin and evaluate capital in non-life insurance applications. By utilizing the entire range of conditional quantile functions, especially higher quantile levels, we detail how quantile regression is capable of providing an accurate estimation of risk margin. Furthermore, we provide an overview of implied capital based on the historical volatility of a general insurer's loss portfolio using quantile regression. Two modeling frameworks are considered based around parametric and nonparametric quantile regression models which we contrast specifically in this insurance setting.

In the parametric quantile regression context, several models including the flexible generalized beta distribution family, asymmetric Laplace (AL) distribution and power Pareto distribution are considered which we detail how to develop under a Bayesian regression framework. The Bayesian posterior quantile regression models in each case are studied via Markov chain Monte Carlo (MCMC) sampling strategies.

In the nonparametric quantile regression models that we contrast to the parametric Bayesian models we adopted AL distribution as a proxy and together with the parametric AL model, we expressed the solution as a scale mixture of uniform distributions to facilitate implementation. The models are extended to adopt dynamic mean, variance and skewness and applied to analyse two real loss reserve data sets to perform inference and discuss interesting features of quantile regression for risk margin calculations.

Furthermore, we consider the class of recently developed stochastic models that combine claims payments and incurred losses information into a coherent reserving methodology. In particular, we develop a family of hierarchical Bayesian Paid-Incurred-Claims models. In the process we extend the independent log-normal model by incorporating different dependence structures using a Data-Augmented mixture Copula Paid-Incurred claims model.

The utility and influence of incorporating both payment and incurred losses into estimating of the full predictive distribution of the outstanding loss liabilities and the resulting reserves is demonstrated in the following cases: (i) an independent payment (P) data model; (ii) the independent Payment-Incurred Claims (PIC) data model; (iii) a novel dependent lag-year telescoping block diagonal Gaussian Copula PIC data model incorporating conjugacy via transformation; (iv) a novel data-augmented mixture Archimedean copula dependent PIC data model.

Inference in such models is developed via a class of adaptive Markov chain Monte Carlo (MCMC) sampling algorithms. These incorporate a data-augmentation framework utilised to efficiently evaluate the likelihood for the copula based PIC model in the loss reserving triangles.

Declaration

I declare that this thesis represents my own work, except where due acknowledgement is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualifications.

Signed ____

Xiaodan (Alice) Dong

Acknowledgements

First and foremost, I would like to express my special appreciation and thanks to my supervisors, Dr. Jennifer S. K. Chan and Dr. Gareth W. Peters for supporting me in the past six years. Their rich intellect, inspiration, enormous guidance and enthusiasm in research have provided a good basis for the present thesis. Both of you have been tremendous mentors for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice on both research as well as on my career have been priceless. Moreover, it is a pleasure to thank Prof. Neville Weber and Associated Prof. Shelton Peiris, who offer me the opportunity to start my research degree with solid foundations and all the friends and staff in the School for their cares and support. Without any of them, I would not have such wonderful experiences.

Above all, I am indebted to my family for their unconditional love and support, and I would also like to thank God for the wisdom and perseverance that he has been bestowed upon me during this research project, and indeed, throughout my life. Their endless love and support have been invaluable in helping me to focus on my academic pursuits. Without them, I would not have had the courage to complete this thesis.

CONTENTS

List of Symbols

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1. Loss reserving background

Loss reserving in general insurance is a vital activity to the success of the insurer. Its basic purpose is to estimate the cost of all future claims arising from policies currently in force and policies written in the past. This uncertain cost is usually the most important figure on its financial statement. Before embarking on the methods and techniques for claims reserving, which make up the main part of the thesis, it is important to understand the data. In the following paragraphs, I will start with outlining the data format of the loss reserving models.

In the area of non-life insurance reserving, there are primarily two types of data used: aggregated claims triangle data and individual data. Individual loss data refers to data on a claim level. Claims in non-life insurance are triggered by an accident which is an event that causes damages covered by an insurance contract. The year of claims occurrence is called the accident year. Typically, claims will not be settled immediately due to various factors, such as investigations and administration process. The delay in years is reported as development year. For each open claim, case officers in insurance companies assign an estimate of future payments in respect of individual reported claims. This estimate is often refereed to as case reserves. Individual claims data is commonly summarized into a aggregated claims triangle, grouping data into rows of accident years and columns of development years. The claims triangle reflects the change in amounts as claims mature. However, the downside is that grouping data lose information as many claims data points will be grouped into one data point per development period per origin period, which means that we will have limited information to get a real sense of the distribution. An example of claims triangle are reported in Figure 1.1, which are the paid out claim amounts Y_{ij} for an insurance company.

They are summarized by accident years (i) and development years (j) covering periods from 1 to n. Apart from the paid claims triangle, there is frequently a triangle of incurred data, which is the quantity obtained by adding the case reserves to the paid claims, and is often available. The second data format which includes these two types of triangles, payment and incurred data together, are illustrated in Figure 1.2. Chapter two of this thesis considers reserving models for both aggregated and individual data. Chapter three examines quantiles for aggregated data. Both chapters consider data involving only one run-off triangle of claim payments. Then chapter four puts forward a model which consider both of the payment and incurred data in two run-off triangles.

Apart from payment and incurred data, there are some other data available for reserving purpose. Generally, the number of claims reported, the average claim size and case reserves data can be summarized in a claims triangle format. Case reserves data is commonly used in the Projected Case Estimates (PCE) model whose performance is usually better in the earlier accident years than payment data based models. It is due to the fact that for more mature claims, the case reserves data is generally more reliable as more information has been collected as claims mature. Using both data formats, we can project an estimate for the ultimate claims amount and therefore the reserves that an insurance company should held by modeling the relationship between successive data across accident and development years. This approach has proved successful and popular. The common industry practice is that for the earlier accident years, PCE model is adopted because of the reliable case estimates.

		Development year j							
		1	2	3	.	.	.	n-1	n
Accident year i	1	$Y_{1,1}$	$Y_{1,2}$	$Y_{1,3}$	.	.	.	$Y_{1,n-1}$	$Y_{1,n}$
	2	$Y_{2,1}$	$Y_{2,2}$	.	.	.	.	$Y_{2,n-1}$	
	3	$Y_{3,1}$	.	.	.	.	.		
	.	.	.	.	.	.			
	.	.	.	.	.				
	.	.	.	.					
	n-1	$Y_{n-1,1}$	$Y_{n-1,2}$						
	n	$Y_{n,1}$							

FIGURE 1.1. Run Off Triangle

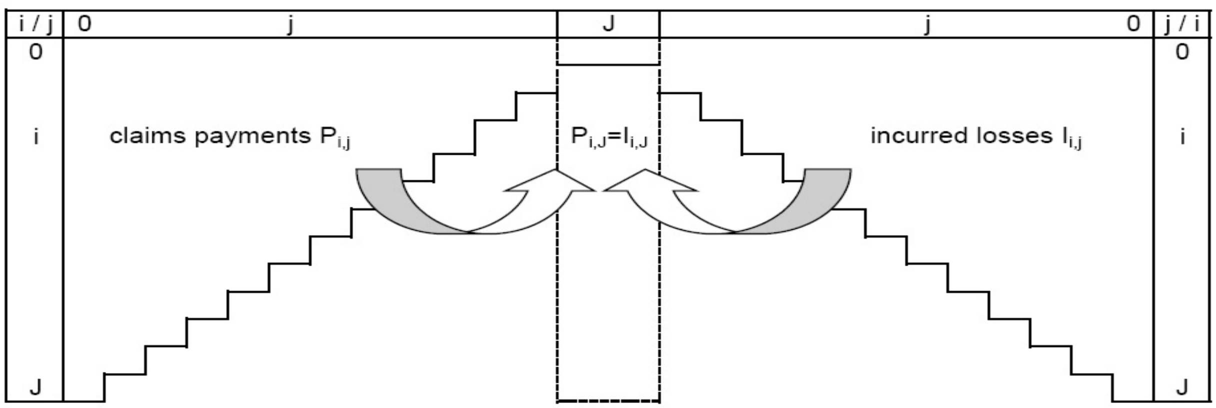


FIGURE 1.2. Claims triangle for payment data and incurred data (source Merz and Wuthrich. (2010)).

1.2. Overview of loss reserving models

Loss reserve analysis has a long and rich history. Numerous approaches to estimate the necessary reserve provisions have been developed to give reasonable estimates. They are usually classified as deterministic or non-stochastic and statistical or stochastic (Hossack et al., 1999; Taylor, 2000). The contribution of this thesis is to strengthen the scientific part of the reserving analysis using statistical inference. To this end, the present thesis is dedicated. In the following sections, I will give an overview of the traditional reserving models and the new developments in statistical claims reserving space.

1.2.1. Chain ladder model. In traditional loss reserving methodologies, the chain ladder method is the most widely applied reserving method. Taylor (2000) provides a detailed description of the classical chain ladder method. It has an appealing elegance, with all the inputs and outputs clearly visible, and it gives a clear representation of how claims are expected to develop over time. For analysis involves multiple triangles, n denotes the nth triangle. The algorithmic definition of the chain-ladder model for triangle $n \in \{1, ...N\}$ at time I reads as follows (Merz and Wüthrich, 2007):

1 . Suppose there are constants $f_l^{(n)}(l = 1, ...J-1)$ so that for all i and $j > I - i + 1$:

$$\hat{Y}_{i,j}^{(n)} = Y_{i,I-i+1}^{(n)} \cdot f_{I-i+1}^{(n)} \cdot f_{I-i+2}^{(n)} \cdots f_{j-1}^{(n)}.$$

is an appropriate predictor for $Y_{i,j}^{(n)}$.

2 . The chain ladder factors $\hat{f}_l^{(n)}$ are estimated by:

$$\hat{f}_l^{(n)} = \frac{\sum_{i=1}^{I-l} Y_{i,l+1}^{(n)}}{S_l^{(n)}} = \sum_{i=1}^{I-l} \frac{Y_{i,l}^{(n)}}{S_l^{(n)}} \frac{Y_{i,l+1}^{(n)}}{Y_{i,l}^{(n)}}$$

where

$$S_l^{(n)} = \sum_{i=1}^{I-l} Y_{i,l}^{(n)}.$$

The Chain ladder method is simple, widely used and well understood. However, it has some significant faults. Primarily, it does not include any calendar year effects. More fundamentally, it doesn't use any risk theory and make any assumptions about the way the data have been generated. Nowadays various extensions of the chain ladder model have been proposed. Like any other deterministic and mechanical methods, the chain ladder method can be reinterpreted via a stochastic model by the addition of error terms. Various attempts have been made to extend the chain ladder technique. The most popular ones are the stochastic chain ladder model by Mack

(1993a) which is a distribution-free method using a heuristic parameter estimation method where a distribution free formula for the standard error of the chain-ladder reserve estimate is derived. Recently, Martinez et al. (2012) extends the traditional chain-ladder framework towards the use of extra data sources. It introduces a micro-model of the claims generating process in order to motivate the models used for estimation, which are applied to triangles of aggregated data. Another extension is the overdispersed Poisson model by England and Verrall (2002), which is a bootstrap of the paid chain-ladder model. It assumes observations are independently over-dispersed Poisson distributed and their means are modeled as the product of a row effect and a column effect. The Chain-ladder model has also been extended in the Bayesian context by Gisler (2006). Wĺuthrich et al. (2008) gives some backgrounds on this model, in which a Bayesian chain-ladder approach is presented assuming that the unknown model parameters follow a prior distribution.

1.2.2. The separation method. Another type of popular reserving model is the separation method. It was first introduced by Verbeek. H (1972) in a reinsurance context, and it was developed further by Taylor (1977) to be applicable to the average claim cost. The idea behind the separation method is to distinguish two patterns in the claims data namely, the development pattern for the accident year and calendar year effects, of which inflation is usually the most important. It differs from the chain-ladder in that while the chain-ladder assumes claim proportionality between the development years and projects the inflation present in the past data into the future, the separation method incorporates it into the model underlying the reserving method. When inflation rate is not constant, the separation method is more applicable. Wĺuthrich et al. (2008) provides a comprehensive review on this method. Björkwall et al. (2010) introduces a parametric bootstrap framework within the separation model which enables joint resampling of claim counts and claim amounts.

1.2.3. The payment per claim finalized model. The payment per claim finalized (PPCF) model is commonly used in claims reserving practice. Payment per claim finalized is simply defined as the total claim payments for a given period divided by the number of claims finalized within that period. The PPCF model consists of two submodels which are designed to model the average payments per claim finalized and the probability that a claim finalizes in a particular quarter from two sources of data, namely the claim size and frequency data. The multiplication of the results from these two models, which are the predicted average claims size and claims frequency, gives us the ultimate claim cost prediction. In other words, its projection is based on the product of these two models in a given development period. The model details have been introduced by Fisher and Lange (1973) and Sawkins (1979b).

1.2.4. Overview of statistical claims reserving models. In recent years, a statistical framework for analyzing claims reserving data has been built up, encompasses, extends and consolidates the actuarial methods. McCullagh and Nelder (1989) point out that most of the stochastic models for loss reserving can be formulated by means of a particular family of generalized linear models (GLMs).

1.2.5. Generalized Linear Models. Generalised linear models was first introduced by Nelder and Wedderbu (1972), and form a remarkable synthesis and extension of classical linear models which allow generalization to the mean structure and response probability of distribution. In particular, the mean of a distribution may depend on a linear function of predictors through a link function, and the response probability distribution adopts any flexible distributions. Haberman and Renshaw (1996) gives a comprehensive review of the application of GLMs to actuarial problems, including loss reserving. Verrall (2004) uses a Bayesian parametric model within the framework of GLMs, and also illustrates how they lead to posterior predictive distributions of quantities of interest. They compare outcomes of this approach with results on approximations for the distribution of the discounted loss reserve when the run-off triangle is modeled by a GLM. GLMs are flexible enough to encompass a large class of model extensions, including flexible mean, variance and skewness structures. Chan et al. (2008) applied the ANOVA and threshold mean structure for loss reserve to allow for a change in the development year effect. GLMs can also incorporate different distortional assumptions to describe different data. A wide choice of distributions including the generalized-t (GT)(Chan et al., 2008), Pareto (Zehnwirth, 1994), the Stable family (Paulson and Faris, 1985), the Pearson family (Aiuppa, 1988), the log-gamma and lognormal (Ramlau-Hansen, 1988) and the lognormal and Burr 12 (Cummins et al., 1999), and the generalized Beta of the second kind (GB2) distribution (Cummins et al., 1990) have been applied in loss reserving.

1.2.6. Quantile regression. Although the classical theory of GLMs allows the modeling of the mean of a distribution, it is however, fruitful to go beyond modeling merely the mean of a distribution. Koenker and Basse (1978) introduce the idea of nonparametric quantile regression to estimate the conditional quantile functions. The quantiles of the conditional distribution of the response variable are linked to functions of observed covariates. In this way, the conditional quantile functions provides a more complete view of possible relationships between covariates and response across quantile levels. Instead of minimizing the sum of squared residuals in the mean regression, quantile regression minimizes the sum of absolute errors for the special case of median regression, and minimizes an asymmetrically weighted sum of absolute errors for the rest of conditional quantile functions. Yu and Moyeed (2001) introduces a technique of estimating the parameters of nonparametric quantile regression by employing a likelihood function that is based on the asymmetric Laplace (AL) distribution. The motivation behind this idea is the inclusion of the absolute error term in the likelihood function of AL distribution. Hence, maximising the likelihood of AL distribution is equivalent to miniziming the loss function of the nonparametric quantile regression. Note that this AL proxy distribution is for nonparametric quantile regression with no distribution assumption. Yu and Moyeed (2001) implement this AL model using a Bayesian approach. Hence, the use of the AL distribution provides an effective way for estimating quantile regression models. In other words, AL distribuiton can be used as a proxy distribution to estimate the parameters of the nonparametric quantile regression. Quantile regression has been applied to a wide range of financial and economics problems. Engle and Manganelli (2004) consider the quantile regression for the Value at Risk (VaR) model. Cai (2010) proposes a parametric quantile regression model, the power-Pareto model which provides flexible quantile functions through a combination of quantile functions for both power and Pareto distributions. These combinations enable the modelling of both the main body and tails of a distribution.

1.2.7. Paid incurred models. Apart from the generalization of loss reserve methods to model different quantiles of a distribution, the models can also be applied to more general data. In particular, we can combine two sources of available data, namely the payment triangle and incurred claims triangles. The traditional models which fit either paid or incurred claims data, separately, does not make full use of all the data available, and results in the loss of some information contained in those data. This leads us to construct models for both sources of information, paid and incurred claims in the form of run-off triangles, which allows us to model the dependency between the two run-off triangles. The application of copulas to model two sources of data, as well as their associated dependency has been studied recently. Nelsen (2006a) gives a comprehensive review of the theory of copulas and their use in finance. Tang and Valdez (2005) applies the simulated loss ratios to aggregate losses from different line of businesses using copulas. Jong (2012) introduces a Gaussian copula model to describe dependence between different line of businesses. The concept of copula was first introduced by Sklar (1973) to decompose a n-dimensional distribution function into two parts, the marginal distribution functions and the copula, which describes the dependence part of the distribution. There are three methods of constructing a copula, namely the inversion method, the Geometric method and the algebraic method. More details can be found in Nelson (2006). The popular copula classes in fiance are the archimedean copula families by Genest and MacKay (1986) and the Gaussian copula. Gaussian copula is very popular because of its tractable properties for computation. As a copula captures the dependence relationship in a multivariate distribution, it allows the full specification of the multivariate distribution of random vectors in terms of marginal distributions and dependence structure.

1.3. Bayesian inference and model selection criteria

1.3.1. Bayesian inference. Bayesian methods for claims reserving have been considered in a series of papers by (H Aastrup and Arjas, 1996; Haastrup and Arjas, 1996; Scollnik, 2002; De Alba, 2002; Verrall, 2004; Verrall and England, 2005). The basic idea of Bayesian is summarized as follows: If the random variables $\boldsymbol{x}$ denote claim figures (payment or incurred cost), the Bayes theorem asserts that the posterior distribution for the parameter $\boldsymbol{\theta}$ conditional on data $\boldsymbol{x}$ is proportional to the data likelihood $f(\boldsymbol{x}|\,\boldsymbol{\theta})$ and the prior densities $f(\boldsymbol{\theta})$, which can be expressed as,

$$f(\boldsymbol{\theta}|\,\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\,\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int f(\boldsymbol{x}|\,\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto f(\boldsymbol{x}|\,\boldsymbol{\theta})f(\boldsymbol{\theta}).$$

Available information on the parameters is incorporated through the prior density for $\boldsymbol{\theta}$, which can be modeled to make use of available information or past experience. This is then combined with the likelihood function via Bayes theorem to obtain a posterior distribution for the parameters.

The Bayesian approach constitutes a powerful alternative to nonparametric and classical frequentist methods. The asymptotic behavior of the Bayesian procedure concerns the way in which posterior measures concentrate their mass around a point of convergence. Various studies have proven the consistency and asymptotic normality of posterior distributions using Bayesian method. The Bayesian implementation of loss reserving models has been an area of considerable interest. First, they allow actuaries to formally incorporate expert or existing prior information. Even when prior information is not available, we can use non-informative or reference priors. Secondly, Bayesian methods can obtain the complete probability distribution for the quantities of interest, either the parameters, or the future values of a random variable. In claims reserving practice, quite often, not only the mean is of interest, but also other quantities, such as quantiles for value of risk measure. In this regard, the Bayesian method provides the an adequate understanding and information of the complete distribution.

1.3.2. Gibbs sampling. As the posterior distribution might not have a non-standard form, Gibbs sampling (Smith and Roberts, 1993) is a commonly used statistical Markov chain Monte

Carlo (MCMC) simulation method to draw samples from the posterior conditional distributions of model parameters. Theoretically, the probability distribution of a parameter given all the other parameters and the past claims data is obtained for each parameter, and then samples are simulated from a Markov chain that has its stationary distribution equal to the joint probability distribution of the parameters and data. Gibbs sampling allows people to draw one parameter or a small block of parameters at a time conditional on values of the other parameters, which is usually much easier than drawing from the posterior for all parameters simultaneously. The idea of the Gibbs sampling is summarized as follows. Assume that we have two model parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$. The joint posterior distribution is written as $f(\theta_1, \theta_2|\boldsymbol{x})$ and the conditional density of one parameter given the other two parameters are written as $f(\theta_1|\,\theta_2, \boldsymbol{x})$ and $f(\theta_2|\,\theta_1, \boldsymbol{x})$ respectively. The algorithm for the implementation is illustrated below:

1. Begin at starting values of $\theta_1^{(0)}$ and $\theta_2^{(0)}$.
2. Draw $\theta_1^{(1)}$ from the conditional distribution $f(\theta_1|\,\theta_2^{(0)}, \boldsymbol{x})$.
3. Draw $\theta_2^{(1)}$ from the conditional distribution $f(\theta_2|\,\theta_1^{(1)}\boldsymbol{x})$ using the newly simulated $\theta_1^{(1)}$.
4. Repeat Step 2 to 3 until R iterations have completed with the simulated values converged to the joint posterior density function.

1.3.3. Metropolis Hastings. Apart from Gibs sampling, the Metropolis Hastings (MH) algorithm by Hastings (1970); Metropolis et al. (1953) has been used extensively in Bayesian inference, to sample from complicated high-dimensional distributions. The MH algorithm in algorithmic form initialized with the arbitrary value $\theta^{(0)}$ is summarized below:

1. Given the current value $\theta = \theta^{(k)}$ at iteration k, sample a candidate parameter $\theta^{'}$ from the candidate-generating density $q(\theta^{'}|\theta)$. with probability α, set $\theta^{(k+1)} = \theta^{'}$, where

$$\alpha = \min(1, \frac{f(\theta^{'}|x)q(\theta|\theta^{'})}{f(\theta|x)q(\theta^{'}|\theta)}) \tag{1.1}$$

2. else set $\theta^{(k+1)} = \theta^{(k)}$.
3. Return the values $\theta^{(k+1)}\{\theta^{(1)}, \theta^{(2)}..., \theta^{(N)}\}$.

4. Repeat for $k = 1, 2, ...N$ to result in a posterior sample.

Chib and Greenberg (1995) provide a detailed review of the more general MH algorithm as well as an explanation of how and why it works. Under Chib and Greenberg (1995)'s more general definition, there are two applications of the MH algorithm, one for implementing acceptance-rejection sampling when a blanketing function is not available and the other for implementing the algorithm with block-at-a-time scans. In the latter situation, Gibbs sampling is a special case of the MH algorithm. The MH algorithm is particularly useful in the context of nonstandard posterior distributions because the normalizing constants for the posterior distributions need not be calculated (Chib and Greenberg, 1995). In some cases when the posterior distribution is not in a closed or standard form, the MH algorithm is applicable, as it is not limited any distribution. When the resulting model is too high-dimensional for MH algorithm, Metropolis-within-Gibbs algorithm (Metropolis et al., 1953) is appropriate. Another example of MH is the random walk MH. The basic idea is as follows. Firstly, we consider each variable in turn. For each variable, we propose updating its value by adding a $N(0, \sigma^2)$ increment. That proposal is then accepted or rejected according to the usual Metropolis ratio. This process is repeated many times, allowing all the variables to converge.

1.3.4. Adaptive MCMC. Although the MH algorithm has been widely used, the tuning of associated parameters ,such as proposal variances σ^2, can be very difficult. Adaptive MCMC algorithms is designed to deal with this problem to achieve efficient mixing at some desirable acceptance rates, say around twenty to thirty percent. The algorithm samples parameters via a learning procedure where the transition kernel specifies the probability to move from one state to another. The transition kernel of the algorithm is sequentially tuned during the simulation in order to obtain optimal efficiency (see Gilks et al. (1998); Haario et al. (2001); Andrieu and Moulines (2006a)). One way to measure the efficiency of MCMC is the rate of convergence to the stationary distribution and the speed of mixing in MCMC algorithms. If the mixing is effective, there should be no homogenous Markov chains. The random walk updating rules with adaptive MCMC can improve the mixing efficiency. The general framework of adaptive MCMC is summarized as follows:

1. Define a measurable function $g_n \times \theta^{k+1} \to \theta$ for $k = 1, 2, ...N$, where $g(\theta|\gamma)$ is the transition kernel.
2. Initialized the adaption chain with some arbitrary but fixed values $(\gamma_0, \theta_0) \in \Gamma \times \Theta$.
3. At iteration $k \geq 1$, given $(\gamma_0, \theta_0, ...\theta_{k-1})$ and $\gamma_{k-1} = g_{k-1}(\gamma_0, \theta_0, ...\theta_{k-1})$ with the convention of $g_0(\gamma, \theta) = \gamma$.
4. Return the value of θ_k according to the transition probability of $P_{\gamma_{k-1}}(theta_{k-1}, .)$ and $\gamma_k = g_k(\gamma_0, \theta_0, ..\theta_k)$.

In Haario et al. (2005b) an adaptive Metropolis algorithm with proposal covariance adapted to the history of the Markov chain was developed. In Andrieu and Thoms. (2008) a tutorial discussion of the proof of ergodicity of adaptive MCMC under two simpler conditions known as *Diminishing Adaptation* and *Bounded Convergence* is presented. *Diminishing Adaptation* means that the total variation of parameter learning at the beginning can be large, but eventually will diminish. The condition of diminishing adaptation is fulfilled when the amount of adaptation diminishes with the length of the chain. *Bounded Convergence* implies that if we take the transition kennel from any pint in the learning process and look at each point, eventually there should have some bound. The law of large number will apply, and therefore make sense to take average as a reasonable estimate. A Markov chain is ergodic if there is a strictly positive probability to pass from any state to any other state in one step. We note that when using inhomogeneous Markov kernels it is particularly important to ensure the generated Markov chain is ergodic, with the appropriate stationary distribution. Two conditions ensuring ergodicity of adaptive MCMC are known as *Diminishing Adaptation* and *Bounded Convergence*. These two conditions are summarised by the following two results for generic Adaptive MCMC strategies on a parameter vector $\boldsymbol{\theta}$. As in Roberts and Rosenthal. (2009) we assume that each fixed MCMC kernel Q_γ, in the sequence of adaptions, has stationary distribution $P(\cdot)$ which corresponds to the marginal posterior of the static parameters.

Define the convergence time for kernel Q_γ when starting from a state in the parameter space E, $\boldsymbol{\theta} \in E$, as $M_\epsilon(\boldsymbol{\theta}, \gamma) = \inf\{s \geq 1 : \|Q_\gamma^s(\boldsymbol{\theta}; \cdot) - P(\cdot)\| \leq \epsilon$. Under these assumptions, they give the following two conditions which are sufficient to guarantee that the sampler produces draws from the posterior distribution as the number of iterates tend to infinity. The two sufficient conditions are:

- *Diminishing Adaptation:* $\lim_{n\to\infty}\sup_{\boldsymbol{\theta}\in E}\|Q_{\Gamma_{s+1}}(\boldsymbol{\theta}, \cdot) - Q_{\Gamma_s}(\boldsymbol{\theta}, \cdot)\| = 0$ in probability. Note, Γ_s are random indices.
- *Bounded Convergence:* For $\epsilon > 0$, the sequence $\{M_\epsilon(\boldsymbol{\theta}, \Gamma_j)\}_{j=0}^{\infty}$ is bounded in probability.

The sampler converges asymptotically in two senses,

- *Asymptotic convergence:* $\lim_{j\to\infty}\|\mathcal{L}\text{aw}([\boldsymbol{\theta}](j)) - P(\boldsymbol{\theta})\| = 0$
- *Weak Law of Large Numbers*: $\lim_{j\to\infty}\frac{1}{j}\sum_{i=1}^{j}\phi([\boldsymbol{\theta}](i)) = \int\phi(\boldsymbol{\theta})P(d\boldsymbol{\theta})$ for all bounded $\phi : E \to R$.

In general it is non-trivial to develop adaption schemes which can be verified to satisfy these two conditions. In this chapter we use the adaptive MCMC algorithm to learn the proposal distribution for the static parameters in our posterior $\boldsymbol{\Phi}$. In particular we work with an Adaptive Metropolis algorithm utilizing a mixture proposal kernel known to satisfy these two ergodicity conditions for unbounded state spaces and general classes of target posterior distribution, see Roberts and Rosenthal. (2009) for details.

The software `WinBUGS` is exploited in this thesis. It carrys out MCMC simulation using Gibbs sampling, which reduces the complexity of sampling from the high-dimensional posterior distribution. It has both a user friendly graphical interface as well as a programming interface for more sophisticated modeling. It contains several useful built-in functions, including the burn-in, thin and autocorrelation function (ACF). The burn-in period function enables us to discard the first B iterations as the burn-in period to ensure that convergence has reached. In this thesis, we discard the first 10,000 integrations. From the remaining $(R - 10,000)$ iterations, parameters are sub-sampled or thinned from every $H^{th} = 50$ iteration to reduce the autocorrelation in the samples. Resulting samples will consist of $M = \dfrac{R - 10,000}{50}$ realizations with their mean taken as the parameter estimates. After that, the history and autocorrelation function (ACF) plots are checked to ensure the convergence and independence of the posterior sample.

1.3.5. Model selection criteria. To assess the model-fit, three criteria: the R percentage, deviance information criteria (DIC) and Bayes factors are adopted. The R percentage is the mean of predicted over actual loss less one, which is a popular measure to quantify the difference between actual and predicted values whereas DIC originated by (Spiegelhalter et al., 2002) is a Bayesian analogue of Akaike's Information Criterion (AIC) which is commonly used in Bayesian analysis. DIC consists of a measure of model fit which is the posterior mean deviance, and a measure of model complexity which is an estimate of the effective number of parameters. It has a competitive advantage over the traditional AIC as it is not only limited to nested models.

The DIC is given by

$$DIC = -\frac{4}{M}\sum_{m=1}^{M}\sum_{i=1}^{I}\sum_{j=1}^{I+1-i}\ln\left[f(y_{ij}|\boldsymbol{\theta}^{(m)})\right] + 2\sum_{i=1}^{I}\sum_{j=1}^{I+1-i}\ln\left[f(y_{ij}|\bar{\boldsymbol{\theta}})\right] \tag{1.2}$$

where $\boldsymbol{\theta}^{(m)}$ denotes the vector of parameter estimates in the m-th iteration of the posterior sample M, $\bar{\boldsymbol{\theta}}$ denotes the posterior mean of $\boldsymbol{\theta}^{(m)}$ and $f(y_{ij}|\boldsymbol{\theta})$ represents the observed likelihood for each observation.

Bayes factors (Kass and Raftery, 1995) are commonly used for pairwise comparisons between models in Bayesian applications. Assuming two models are regarded as equally probable a priori, a Bayes factor represents the ratio of the posterior probabilities of the models. The model which is a posteriori most probable is determined by whether the Bayes factor is less than or greater than one. However, for some certain complex models, it might be difficult to computer Bayes factor accurately. In these cases, the Bayesian information criterion (BIC) gives rough approximation to the logarithm of the Bayes factor, which is easy to use and dose not require the evaluation of the prior distributions. (Kass and Raftery, 1995). In this thesis, we use the R percentage and DIC for Chapter two and three, the Bayes factor for chapter four.

1.4. General insurance loss reserving problems

The claims reserving problem has been studied rather extensively. Long tail business in general insurance is characterized by lengthy delay between the period of cover and either the emergence or settlement of claims. The critical problem with setting reserve for long-tail casualty classes of business is that the estimation of ultimate net losses and loss expenses is a complex process due to a number of factors. Firstly, data may takes years to develop or report, allowing claims to be exposed to more legislation changes or other unexpected events. Hence it is necessary to consider a dynamic mean function and heavy tailed data distribution. Secondly, a relatively low proportion of net losses would be reported claims and expenses and an even smaller percentage would be net losses paid. Therefore, incurred but not reported (IBNR) would constitute a relatively high proportion of net losses which constitutes a large degree of uncertainly for setting reserves.

A variety of methods are employed to estimate losses for long-tail casualty classes of businesses. These methods ordinarily involve the use of loss trend factors intended to reflect the annual growth in loss costs from one accident year to the next. They perform poorly on longer tailed line of business due to the additional assumptions needed for development factors of later development periods. The additional assumptions are, for instance, legislative changes and unexpected events during the long pay out periods. Recently, some dynamic models are developed within the GLMs framework which involves flexible distribution such as the ones mentioned in Section 1.1. We propose the use of the GB2 distribution in Chapter 2 to predict loss reserve. The GB2 family provides flexible tail estimates, and therefore can be used to model heterogeneous loss reserve data without doing repetitive distribution testings. It includes both heavy-tailed and light-tailed severity distributions which is useful to describe different loss reserve data. Furthermore, in analyzing aggregate loss with long tail lines, it is very likely that practical issues arising in reality, such as legislative changes during the long lag period of claims exposure, will affect claim payments. Failure to account for these factors will result in severe bias in loss reserve. We extend the mean of the GB2 distribution to adopt some dynamic models including the ANOVA, state space and threshold models. It is shown through both simulation study and forecasting that these features increase prediction accuracy.

Some volatile lines of business exhibit the pattern of either a big loss or no loss. It usually presents when a portfolio has a small market share or portfolios, such as the homeowners insurance when the hurricane seasons come. Setting reserves for these kind of portfolios is not easy. The problem arises when the observed development figures within a given loss development triangle heavily fluctuate due to random fluctuations and a scarce data base. This makes it difficult to make a reliable forecast for the ultimate claim. In such situations, actuaries often rely on industry-wide development patterns rather than on the observed individual data. Moreover, setting appropriate risk measures for the total loss distribution of the institution, such as the Value-at-Risk and the Expected Shortfall risk measures, rely upon the accurate specification of a tail functional of the total loss distribution, which is typically a very extreme quantile level. Whilst the calculation of this quantile level can be performed in a number of different ways, it is the intention of this thesis to provide actuaries and risk managers with a mathematically rigorous framework to understand the behavior of these risk measures with respect to measurable factors that directly drive the key loss processes. In Chapter 3, We describe how quantize regression and variance structures can be used to solve this problem. By utilizing the entire range of conditional quantile functions, especially higher quantile levels, we detail how quantile regression is capable of providing an accurate estimation of risk margin. An overview of implied capital based on the historical volatility of a general insurers loss portfolio is presented using the best quantile regression model.

Traditionally, actuaries that adopt a stochastic framework would evaluate claims liability using a central estimate which is typically defined as the expected value over the entire range of outcomes. However with the inherent uncertainty that may arise from such an estimator which is not statistically robust and therefore sensitive to outlier claims, claims liability measures often differ from their central estimates. In practice, the approach adopted is typically to then set an insurance provision so that, to a specified probability, the provision will eventually be sufficient to cover the run-off claims. Risk margin is defined as the amount required to ensure the value of the technical provisions is increased from the discounted best estimate to this probability of sufficiency or adequacy of liabilities. There are two commonly used methods for risk margin estimation: the cost of capital and the quantile methods.

1.4.0.1. *cost of capital method.* Cost of capital approaches describe the risk margin as equivalent to the cost required to set up and maintain capital to support the risks to which a firm is exposed. Such methods require projecting the relevant capital required to support risks forward for the lifetime of the existing business, and then multiplying this by the cost to the firm of raising this capital, before discounting to get a present value (Brown, 2012). However, the Cost of Capital method fails to satisfy a number of desirable properties particularly for liabilities with very long maturities. For example it has no upper bound related to the Capital Requirement or the maximum value of liability, and it is not invariant under the choice of time unit(Waszink, 2013).

1.4.0.2. *Quantile method.* The second method, which is the Quantile or percentile method, was first described for regulatory purpose by the Australian regulator (APR) in the prudentiaa standard GPS 210 - Liability Valuation for General Insurers. Quantile methods describe the risk margin as the difference between liabilities valued at a set percentile and at their best estimate. Opinions vary as to what level the percentile should be set, and it should be noted that for liabilities with highly skewed (fat-tailed) risks, even a relatively high percentile may result in the value including risk margin of liabilities being lower than the best estimate (Brown, 2012). It is worth noting that the more volatile a portfolios runoffs or those that display heavy tailed features may require a higher risk margin, since the potential for large swings in reserves is greater than that of a more stable portfolio. Under the percentile method, insurers typically set a provision to be sufficient to cover the run-off claims at a 75 percent probability level. Compared to mean model, quantile regression model is more statistically robust and can detect the dependence on the covariates in the lower and upper tails of the response distribution. We are intended to introduce the notion of quantile regressions to be incorporated into reserving and capital measurement. What we bring to the quantile based framework in our proposed methods in Chapter 3 is the ability to incorporate in a rigorous statistical manner, regression factors based quantile regression framework with skewness function to estimate risk margin explicitly. We adopted an asymmetric Laplace (AL) distribution as a proxy and the model is extended to adopt dynamic mean, variance and skewness functions.

However a univariate quantile regression model is incapable of addressing the cross correlation of related policy procedures and hence will underestimate their volatilities and subsequently their future claims liabilities. Considering multiple sources of data allows actuaries to best utilize the available information for loss reserves and improve prediction accuracy. The Munich chain ladder method introduced by Quarg and Mack (2004) is one of the first claims reserving approaches in the actuarial literature to unify outstanding loss liability prediction based on both sources of information. This method aims to reduce the gap between the two chain ladder predictions that are based on claims payments and incurred losses data, respectively. The main drawback with the Munich chain ladder method is that it involves several parameter estimates whose precisions are difficult to quantify within a stochastic model framework. Merz and Wuthrich. (2010) introduced a log-normal PIC chain model and used Bayesian methods to estimate the future part of the claims reserving triangles based on both payment and loss incurred information. Its major advantage is that the full predictive distribution of the outstanding loss liabilities can be quantified. One important limitation of the model of Merz and Wuthrich. (2010) is that it does not develop the dependence properties of the PIC model that will be applicable to loss reserving data observed in practice. In Chapter 4, we extend the proposed Bayesian PIC chain-ladder models of Merz and Wuthrich. (2010) to capture additional dependence structures as it is well known that the dependence within payment data, within incurred loss data, and between payment and incurred loss data commonly exists due to the nature of the loss process.

1.5. Objective and structure of the thesis

The remaining thesis is structured as follows. Extensions of the GB2 model and mixture of GB2 model using different mean functions will be investigated in Chapters 2. Quantile function of the GB2 model and other parametric and non-parametric quantile regression model for risk margin estimation will be discussed in Chapter 3, Chapter 4 combines two sources of data and discuss bivariate model with copula function. Lastly, Chapter 5 will summarize this research with some concluding remarks and implications for future developments.

CHAPTER 2

Loss Reserving Using Dynamic Structure and Generalized Beta Distribution

This Chapter presents the modeling of long tail loss reserving data using the generalized beta distribution of the second kind (GB2) with dynamic mean functions and mixture model representation.

2.1. Background

In order to ensure an insurance company's financial security, it is necessary to estimate future claims liabilities. Reserving for the amount of future claims payments involves a large degree of uncertainty, especially for long tail class business where tail behaviors can be largely different. Hence it can be difficult to estimate the loss reserve precisely. Traditionally, conventional distributions such as the lognormal and gamma are used to model severity (Taylor, 2000). In making these distributional assumptions, researchers may underestimate the risk inherited in the long tail which is affected by large claim liabilities because these distributions do not possess flexible tails to describe the features of large claims. Failure to estimate the large claim liabilities adequately can cause financial instability of the company and eventually lead to insolvency. In order to improve modeling accuracy and reliability, sophisticated loss models have been derived with different distributional assumptions.

A wide choice of distributions including the generalized-t (GT) (Chan et al., 2008), Pareto (Zehnwirth, 1994), the Stable family (Paulson and Faris, 1985), the Pearson family (Aiuppa, 1988), the log-gamma and lognormal (Ramlau-Hansen, 1988) and the lognormal and Burr 12 (Cummins et al., 1999), have been used in loss reserving. While the GT distribution is flexible and nests several important families of distributions including the Student-t, uniform, both leptokurtic and platykurtic, and exponential power, it requires log-transformation for the loss data and the resulting log-linear model is more sensitive to low values than large values. As the residuals are negatively skewed when the data contain low claims, Chan et al. (2008) suggested skewed heavy-tailed distributions such as the skewed-t distribution. This chapter remedies the drawback of GT distribution and proposes the flexible generalized beta distribution of the second kind (GB2) to model severity distribution. The GB2 family provides flexible tail estimates, and therefore can be used to model heterogeneous loss reserve data without doing repetitive distribution testings. It includes both heavy-tailed and light-tailed severity distributions, such as The Gamma, Weibull, Pareto, Burr12, lognormal and the Pearson family, hence providing convenient functional forms to model insurance claims (Cummins et al., 1990, 2007).

In traditional methodology of estimating severity distribution, loss data is summarized by accident and development periods, thereby adopting a single aggregated loss distribution (Mack, 1991; Chan et al., 2008). However, the question arises as to whether estimation should be carried out on aggregate loss, or individual loss, where individual losses are observed but the interest is focussed on the sum. Estimating loss reserve using aggregated data reduces the impact of potential outliers (Chan *et al.*, 2008); however, in the process of aggregating data, individual information and variability are lost. Recently, more analyses are performed on individual claims (Taylor and McGuire, 2004). Cummins et al. (2007) considered subgroups of claims by accident and development years, and he applied separate GB2 distribution with a constant mean to each cell of the runoff triangle. This model allows greater flexibility than fitting a single severity distribution across cells. However, as a separate model is fitted to each cell, the model does not consider covariate effects, and hence the trend movement of claims across development years cannot be obtained in a statistically efficient manner. In order to capture the effect of individual characteristics, accident and development years, a single model with dynamic distributional parameters should be fitted to the entire set of individual data. Indeed, data of these two types, namely aggregated or individual, which possess very different characteristics should be handled separately. This chapter proposes two types of loss reserve models applied specifically to aggregated and individual claims.

In analyzing aggregate loss with long tail lines, it is very likely that practical issues arising in reality, such as legislative changes during the long lag period of claims exposure, will affect claim payments. Failure to account for these factors will result in severe bias in loss reserve. In order to cater for these irregular claims behaviors, we extend the mean of the GB2 distribution to adopt some dynamic models including the ANOVA, state space and threshold models. Threshold models, first introduced in Tong (1978), can be considered when the behavior predicted by the model differs in some important ways, for example, a shift in the mean and/or distributional parameters when a switching variable, such as the accident year exceeds certain thresholds, thus offering a dynamic modelling mechanism of risk factors across a threshold. They are characterised by easy interpretation and consequently a large number of applications to real phenomena can be found. Among some of these applications, Li and Lam (1995); Ling (1999) investigated the asymmetry and the volatility of financial markets; Montgomery et al. (1998); Koop and Potter (1999) used these structures to model unemployment rate. Chan et al. (2008) applied the threshold model for loss reserve to allow for a change in the development year effect.

Many dynamic models can usefully be written in a state-space form. The properties of state-space model can be found in Hamilton (1994). Verrall (1989) proposed a state space representation of the chain ladder linear model, De Jong and Penzer (2004) present the ARIMA model in state space form, and Chan et al. (2008) used the state space form to model loss reserve data. While the ANOVA model enables the effects of accident years and lag years to act separately on the total loss, the state space model is a dynamic modelling approach which allows parameters to evolve in a flexible time-recursive manner, and thereby allowing interaction effect between accident and development years. It provides a flexible and unified framework to specific and often complicated circumstances (De Jong and Zehnwirth, 1983).

To analyse individual loss data, controlling for unobserved heterogeneity is an important issue. McDonald and Butler (1987) demonstrated how mixture distributions can be applied to model heterogeneous data. We propose the mixture presentation for GB2 distribution, which allows the dynamic mean and shape parameters to vary across subgroups of claims. The estimated group membership for each observation enables classification of claims into different risk groups. Such information is useful for the managers of insurance companies to derive separate strategies for handling different subgroups of claims with varying risk characteristics.

For model implementation, the use of Bayesian ideas and techniques for loss reserving, dates back to 2000 when Verrall (2000) utilized the Bayesian approach to forecast outstanding claims payments in the lower runoff triangle. The benefit of using Bayesian procedure, lies in the adoption of available prior information and the provision of a complete predictive distribution for the required reserves (De Alba, 2002). Different Bayesian loss reserve models have been proposed for different types of claims data. Zhang et al. (2012) proposed a Bayesian non linear hierarchical model with growth curves to model the loss development process, using data from individual companies forming various cohorts of claims. This model allows pooling of information from multiple companies to perform cross-company analyses. Ntzoufras and Dellaportas (2002) investigated various models for outstanding claims problems using a Bayesian approach via Markov chain Monte Carlo (MCMC) sampling strategy and showed that the computational flexibility of a Bayesian approach facilitated the implementation of complex models, such as the state space and threshold models. We adopt the Bayesian approach and propose alternative hierarchical forms of the models through the scales mixtures representation of the GB2 distribution. This approach substantially simplifies the Gibbs sampler without a heavy computational cost.

Although GB2 distribution provides flexible tails for modeling loss data, its usage is still very limited, particularly in analyzing aggregate loss under the Bayesian framework. To enable accurate loss reserving, objectives of this chapter are two-fold: to derive unique modeling strategies to analyze two types of loss data, the aggregate loss in the runoff triangle and the individual loss, which possess very different features and characteristics, using the flexible GB2 distribution with dynamic mean functions and to implement the proposed models using Bayesian approach.

The rest of this chapter is organized as follows: In section 2.2, we introduce the GB2 distribution, with its properties and its relationship with other distributions. This chapter forms the paper published as Dong and Chan (2013). Section 2.3 describes the Bayesian approach and how it is incorporated in our models. We present our empirical study, with different modelling strategies to address the practical issues arising in aggregated and individual loss reserving data, in sections 2.4 and 2.5. Section 2.6 provides some final remarks and the conclusion from this study.

2.2. The Generalized Beta Distribution

Loss data often exhibits heavy-tailed behavior, particularly for long tail business class. The generalized beta distribution of the second kind (GB2) has attractive features for modelling loss reserve data, as it nests a number of important distributions as its special cases. The GB2 distribution has four parameters, which allows it to be expressed in various flexible densities. The density function is specified as follow:

$$f(y; a, b, p, q) = \frac{|a| y^{ap-1}}{b^{ap} B(p,q)(1 + (y/b)^a)^{p+q}}, \tag{1}$$

for $y > 0$ where b is a scale parameter and a, p and q are the shape parameters such that b, p and q >0 and $a \neq 0$. The beta function $B(p, q)$ is defined by:

$$B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)},$$

and $\Gamma(\cdot)$ denotes the gamma function. Generally, the relative values of p and q determine the skewness of the distribution and negative values of a yield inverse distribution (Cummins et al., 1990). The moments for the GB2 distribution are expressed as follow:

$$E(Y^h) = \frac{b^h B(p + h/a, q - h/a)}{B(p,q)}. \tag{2}$$

In particular, the mean of the distribution is

$$E(Y) = \frac{bB(p+1/a, q-1/a)}{B(p,q)} \tag{3}$$

and the mean and variance of the GB2 distribution exist if and only if $-p < 1/a < q$ and $-p < 2/a < q$ respectively. The density function in (1) can also be expressed as scale mixtures of generalized gamma (GG) distribution as follows:

$$f(y|a,b,p,q) = \int_0^\infty f_{GG}(y|p,\lambda,a) f_{GG}(\lambda|q,b,a)\, d\lambda \tag{4}$$

where λ is the mixing parameter and the density function for the GG distribution is

$$f_{GG}(y|\alpha,\beta,\delta) = \frac{\delta[(\beta y)^\delta]^\alpha}{y\Gamma(\alpha)} \exp[-(\beta y)^\delta]$$

with moments

$$E(Y^k) = \frac{\Gamma(\alpha + k/\delta)}{\beta^k \Gamma(\alpha)}$$

Parameters of the GB2 distribution can be tuned to obtain different special cases and thereby reduce the model complexity. It includes both Pearson and non-Pearson families of distributions. The Pearson distribution first published by Karl Pearson in 1895, is a family of continuous probability distributions, including four types of distributions (numbered I through IV) in addition to the normal distribution. The relationship of GB2 distribution with other distributions is summarized in Figure 2.1 by the distribution tree in Figure 2.1 of Cummins et al. (1990) adopting the part under the GB2 distribution.

Clearly, the GB2 distribution is more general than any other distributions at lower hierarchy of the distribution tree. Figure 2.1 shows that the special cases of GB2 distribution include the 3-parameter distributions of log-t (LT), generalized gamma (GG), beta of type 2 (B2), Burr types 3 and 12 (BR3 and BR12), the 2-parameter distributions of log-Cauchy (LC), lognormal (LN), Weibull (W), gamma (GA), variance ratio (F), Lomax or shifted Pareto (L), Fisk or loglogistic (Fisk) and the 1-parameter distributions of half normal (HN), Rayleigh (R), exponential (EXP), Chi-square (χ^2) and half-t (Ht) distributions.

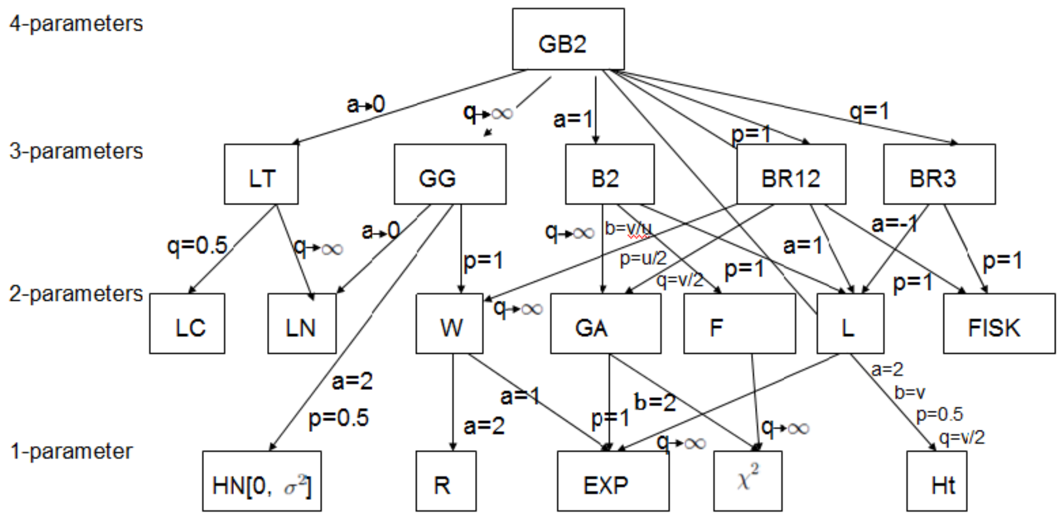


FIGURE 2.1. GB2 Distribution Tree.

Figure 2.2 graphs the probability density functions for some of these special cases whereas Figure 2.3 demonstrates how the density function of the GB2 distribution changes when one or two of the four parameters vary while the others are fixed. The density curves for all special cases are standardized by setting $b = 1$. The density curves for the GB2 distribution and its parameters in Figure 2.3 are highlighted in red. It is clearly shown that as q or p/q decreases, the tail of the distribution becomes heavier. Larger values of a produce heavily right-skewed distributions with thicker tails whereas more negative values of a yield density functions with sharper peaks and longer, fatter tails. By suitably changing the value of a, the GB2 distribution can be expressed in diversified forms ranging from symmetric to heavily right-skewed distributions.

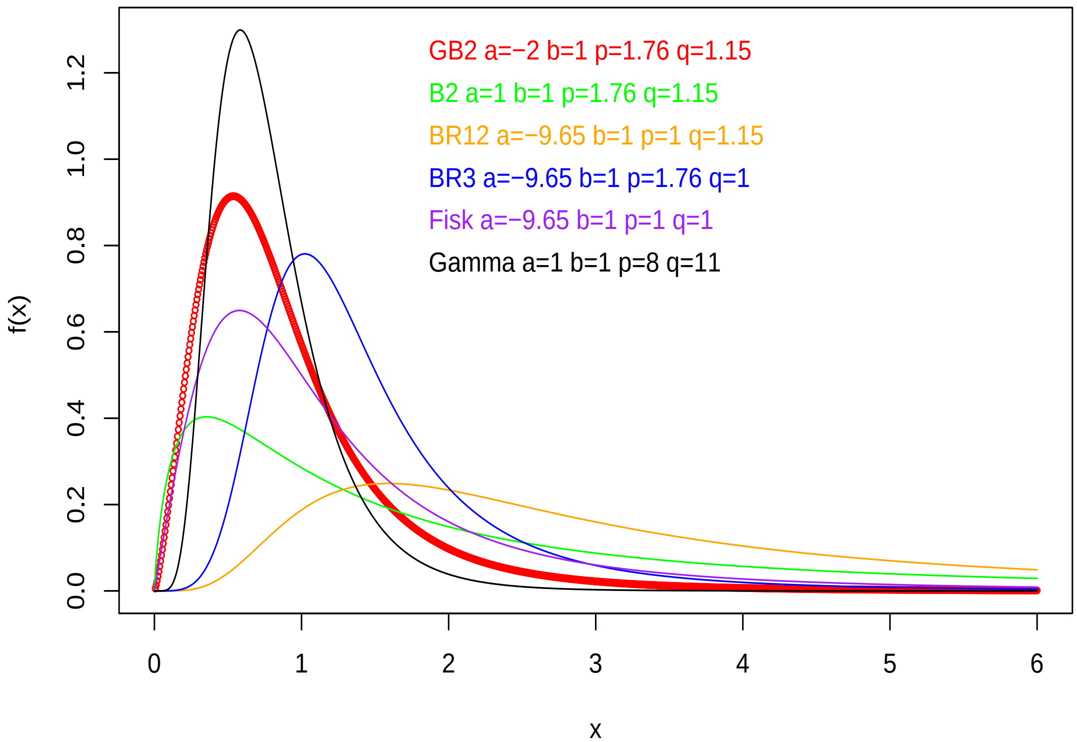


FIGURE 2.2. Probability density function across subgroups of distributions in the GB2 family

2.3. Bayesian methodology

The inclusion of elaborate models, such as the state space, threshold and mixture models into the mean of the four-parameters' GB2 distribution, complicates the likelihood function and its optimization considerably. The Bayesian approach converts the optimization problem in the likelihood approach to a sampling problem, and by making use of the hierarchical structure of the model and MCMC techniques, it lessens the complexity of model implementation for complicated models. In the case of nonstandard posterior distributions, MCMC techniques (Smith and Roberts, 1993; Gilks et al., 1996) with Gibbs sampling and Metropolis Hastings algorithm (Hastings, 1970; Metropolis et al., 1953) produce samples from the intractable posterior distributions of all unknown parameters. Moreover, the prior probability distributions in Bayesian inference provide a powerful mechanism for incorporating information from previous studies, and for controlling confounding.

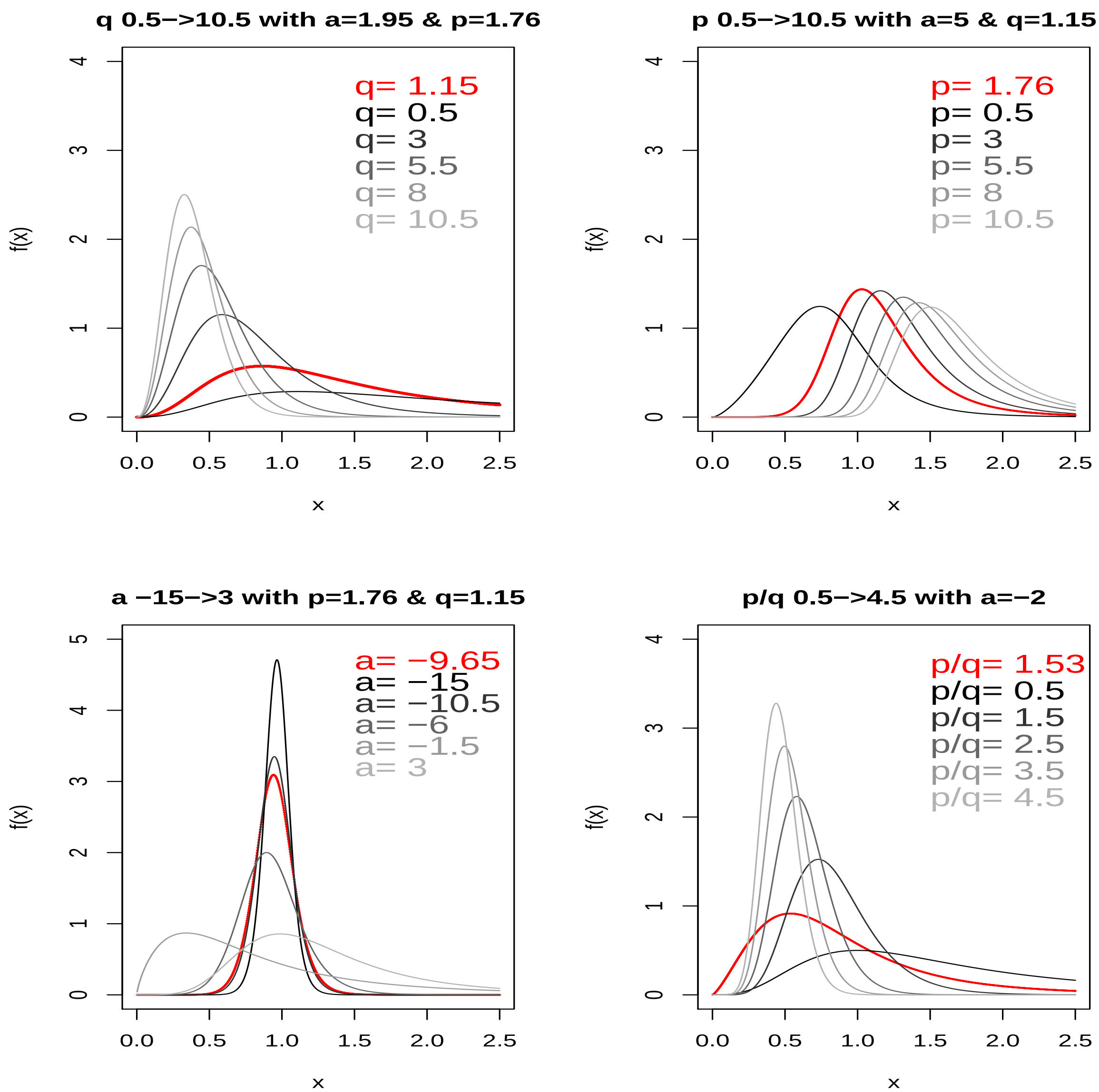


FIGURE 2.3. Probability density function across shape parameters in the GB2 family

Even in the situation where there is no agreement on the prior information, we can use noninformative or reference priors. Inference under this circumstance is so called objective Bayesian inference (Berger, 1985).

Furthermore, the emergence of WinBUGS, a user friendly software for Bayesian inference using MCMC techniques, allows non-experts to perform Bayesian analysis of complex statistical models. In this chapter, all models are implemented via Bayesian approach using WinBUGS and the codes for all models are available upon request. For each model in the empirical study (sections 4 and 5), a single Markov chain is run for 40,000 iterations, discarding the initial 10,000 iterations as the burn-in period and sampling every 30th iteration to mimic a random sample of size 1,000 from the joint posterior distribution for posterior inference. Parameter estimates are given by the posterior means or medians. The autocorrelation functions and history plots are carefully checked to ensure that the posterior samples have converged and are independent. Computation time depends on the complexity of model and power of computer and it is around 4.5 hours using a Core 2 Duo 2GHz PC for fitting the threshold state space models (section 4.2.3) in the empirical study.

2.4. Study of aggregated loss data

Aggregated loss triangles have been widely used to estimate insurance liability. Although sophisticated models have been developed to project the expected payments in the lower runoff triangle, the flexibility of the models to allow for extreme claims and legislative changes during the study period are often uncertain. In this study, we explore the use of the flexible GB2 distribution with four mean functions to allow for some extreme and irregular aggregated losses.

2.4.1. The Data. The data we analyze is the amount of cumulative payments for all the compulsory third party (CTP) policies in Queensland (QLD) as at June 2008. CTP insurance policy covers risk that would be referred to as Auto Bodily Injury in the U.S. and Motor Bodily Injury in the U.K.. In order to remove the influence of inflation for reserving purposes, we utilize the average weekly earning index from the Australian Bureau of Statistics (ABS) to inflate all the values to December 2008 dollars. Hence, the data used in this analysis represents the inflated cumulative payment for QLD CTP portfolio. The data are in the units of millions summarized by accident and development quarters covering periods from December 2002 to June 2008. It contains 276 observations over 23 accident quarters as reported in Figure 2.4.

Since CTP insurance policy covers accidental bodily injury or death of third parties as a result of road traffic accidents, the pay out period is typically long. People who are severely injured in a motor vehicle accident require long term medical treatment and rehabilitation, resulting in

Accumulative Payment	Development Quarter																							
Accident Quarter	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Exposure
Dec-02	0.1	0.7	2.0	4.1	7.1	11.3	17.2	22.8	27.7	34.4	40.6	54.9	67.9	77.7	85.6	94.5	102.5	106.5	115.9	119.3	122.6	124.8	130.2	2.6
Mar-03	0.1	0.7	1.6	2.5	3.8	5.5	8.6	11.2	13.9	19.8	25.9	33.3	41.2	49.7	58.8	68.1	73.5	80.1	88.5	93.8	95.2	97.7		2.6
Jun-03	0.1	0.7	1.4	2.2	3.3	5.0	7.6	10.6	14.9	20.9	27.9	33.7	40.8	51.0	62.0	67.6	75.7	81.9	86.7	88.0	95.3			2.6
Sep-03	0.1	0.7	1.5	2.6	3.9	5.9	8.0	11.6	16.0	24.0	29.8	35.2	46.2	55.8	62.2	72.7	80.1	86.3	91.7	95.8				2.6
Dec-03	0.0	0.6	1.5	2.5	3.5	4.6	7.1	10.8	16.1	20.6	28.0	34.0	46.1	51.4	61.2	69.2	75.9	80.3	88.5					2.7
Mar-04	0.1	0.6	1.6	2.5	3.4	5.1	7.4	11.6	16.6	24.7	32.1	40.8	48.8	58.6	67.3	79.1	89.1	94.7						2.7
Jun-04	0.1	0.6	1.4	2.1	3.1	5.0	7.5	11.6	17.4	23.0	30.2	36.4	46.4	52.9	64.4	71.2	76.6							2.7
Sep-04	0.1	0.6	1.3	2.2	3.3	5.4	8.6	13.4	19.2	26.7	34.7	48.9	58.3	73.2	81.7	93.3								2.8
Dec-04	0.1	0.6	1.4	2.2	3.3	5.1	9.2	15.1	21.6	27.5	39.9	49.2	63.1	69.2	79.2									2.8
Mar-05	0.0	0.5	1.5	2.3	3.4	4.9	9.0	15.7	21.2	31.8	39.7	47.9	56.6	65.3										2.8
Jun-05	0.1	0.7	1.5	2.4	3.7	6.0	11.3	15.7	23.1	31.4	40.5	47.9	63.1											2.9
Sep-05	0.1	0.7	1.8	2.7	5.0	7.9	11.7	19.9	28.8	38.3	45.2	53.8												2.9
Dec-05	0.1	0.8	1.8	3.0	4.5	7.6	13.3	20.4	29.1	36.9	46.8													2.9
Mar-06	0.0	0.5	1.1	2.0	2.9	6.1	10.8	18.0	24.8	31.8														3.0
Jun-06	0.1	0.6	1.4	2.1	4.1	8.5	15.9	22.4	30.1															3.0
Sep-06	0.0	0.6	1.3	2.2	4.0	10.2	17.1	24.8																3.0
Dec-06	0.1	0.5	1.5	3.0	5.0	9.4	17.0																	3.1
Mar-07	0.0	0.7	1.7	2.7	4.3	9.9																		3.1
Jun-07	0.1	0.6	1.4	2.3	4.3																			3.2
Sep-07	0.1	0.8	1.7	2.8																				3.2
Dec-07	0.1	0.6	1.3																					3.2
Mar-08	0.1	0.7																						3.3
Jun-08	0.1																							3.3

FIGURE 2.4. QLD CTP cumulative payment data

substantially high losses. In certain circumstances, if the case goes to court, a legal procedure can be extraordinarily long, and furthermore legal costs can be very high. A substantial part of claims cost is from larger CTP claims which take longer to settle. Hence we propose the GB2 distribution with a flexible tail to model the extreme aggregate loss.

2.4.2. Mean model. In order to capture the irregular claims behaviors, we apply the following four mean models: ANOVA, state space, threshold and state space threshold. For each of the models, we use the accident and development quarter as our covariates, and offset by the number of polices in force (n_i) in each accident quarter.

2.4.2.1. *ANOVA model.* Let Y_{ij} denote the aggregated total claims payment made in accident quarter i, settled in lag quarter j and n_i denote the total number of polices in force in accident quarter i. We apply the two factor ANOVA model (Model 1) as follows:

$$\begin{aligned} Y_{ij} &\sim GB2(a, b_{ij}, p, q), && (5)\\ b_{ij} &= \frac{E(Y_{ij})B(p,q)}{B(p+1/a, q-1/a)} && (6)\\ \log(E(Y_{ij})) &= \mu_{ij} + \ln(n_i), && (7)\\ \mu_{ij} &= \mu_0 + \alpha_i + \beta_j, && (8) \end{aligned}$$

where the parameters α_i and β_j which denote accident quarter and development quarter effects respectively satisfy the following constraints:

$$\alpha_1 = \beta_1 = 0. \tag{9}$$

The following diffuse priors:

$$\mu \sim N(0, 100), \qquad \alpha_i \sim N(0, 100), \qquad \beta_j \sim N(0, 100), \tag{10}$$

$$a \sim N(0, 100), \qquad p \sim Ga(0.001, 0.001), \qquad q \sim Ga(0.001, 0.001) \tag{11}$$

are assigned to the model parameters where $N(\mu, \sigma^2)$ represents the normal distribution with mean μ and variance σ^2; $Ga(r, u)$ represents the Gamma distribution with mean r/u and variance r/u^2. For parameters without restricted ranges, we assign normal distributions with zero mean and large variance as there is no prior information on their values. For shape parameters with a positive range, they are assigned Gamma distributions with unit mean and large variance to reduce the detrimental effect of estimation risk. In general, this set of priors applies to subsequent analyses.

2.4.2.2. *State space model.* The idea behind a state-space representation of a linear model is to capture the dynamics of an observed $(n \times 1)$ vector $\boldsymbol{y}_t$ in terms of a possibly unobserved $(r \times 1)$ vector $\boldsymbol{\xi}_t$ known as the state vector for the model. The dynamics of the state vector are taken to be vector:

$$\boldsymbol{\xi}_{t+m} = \boldsymbol{F^m}\boldsymbol{\xi}_t + \boldsymbol{F^{m-1}}\boldsymbol{v_{t+1}} + \boldsymbol{F^{m-2}}\boldsymbol{v_{t+2}} + \ldots + \boldsymbol{F^1}\boldsymbol{v_{t+m-1}} + \boldsymbol{F^0}\boldsymbol{v_{t+m}} \tag{12}$$

for $m = 1, 2, \ldots,$

where $\boldsymbol{F}$ denotes an $(r \times r)$ matrix and the $(r \times 1)$ vector $\boldsymbol{\upsilon_t}$ is taken to be i.i.d. random vector with zero mean, $\boldsymbol{F^m}$ denotes the matrix $\boldsymbol{F}$ multiplied by itself m times. Hence the optimal m-period-ahead forecast is seen to be

$$E(\boldsymbol{\xi}_{t+m}|\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t+1}...) = \boldsymbol{F}^m \boldsymbol{\xi}_t. \tag{13}$$

Future values of the state vector depend on $(\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, ...)$ only through the current value $\boldsymbol{\xi}_t$. This framework avoids the need for the detailed tailoring of calculations or mode of analysis and hence is suitable in a wide variety of circumstances.

In particular, the state space (SS) model has the flexibility to allow parameters to develop in an auto-regressive process. When the SS model is applied to describe both the accident (α) and j-th development quarter (β_j) effects with $m = 1$ and $r = 1$ in (12), the resultant model (Model 4) is given by (5) to (7) and the following:

$$\begin{aligned} \mu_{ij} &= \mu_0 + \alpha_i + \beta_{ij}, \qquad (14) \\ \alpha_i &= \alpha_{i-1} + \upsilon_i, \\ \beta_{ij} &= \beta_{i-1,j} + \nu_i, \end{aligned}$$

where the interaction between accident and development quarters is incorporated by β_{ij} giving different β_{ij} for different accident and development quarters and α_i and β_{ij} satisfy the following constraints:

$$\alpha_1 = \beta_{1j} = 0. \tag{15}$$

Diffuse priors in (38) and (39) are assigned to the model and further

$$\upsilon_i \sim N(0, 100), \qquad \nu_i \sim N(0, 100).$$

2.4.2.3. *Threshold model.* The threshold (T) model allows the mean function to vary before and after a threshold variable, such as time T. The model is very useful in accounting for events, such as legislation changes and catastrophes in reality. CTP policies normally possess long pay out periods, allowing claims to be exposed to more legislation changes or other unexpected events.

In the aggregated loss data, there is a major legislative change: the Civil Liability Act (CLA) 2003 during the claims accident period. Under this new Act, injuries are assigned a point value between 1 and 100 where zero relates to an injury not severe enough to justify any award of general damages and one hundred is an injury of the gravest conceivable kind (`www.maic.qld.gov.au`).

The main implication is that the mean of total claims payment might shift after the new Act took effect. Figure 2.5 plots the time series of aggregate loss across development quarters for each accident quarter. The time series before and after December 2005 (T=13) are marked in black and red lines respectively. The red lines show sharper increases than the black lines showing a lag effect of CLA 2003. We have tuned the model with T ranging from 12 to 14 and found $T = 13$ provides the best model fit. This can be possibly explained by the fact that the rate of finalisation of claims change considerably across $T = 13$. A further effect is that the claim frequency steadily decreases with increasing accident quarter. This results from the gradual elimination of lower-severity claims and hence the claim sizes displays rats of increase in excess of inflation.

Furthermore, the data set displays a burst of heavy superimposed inflation, which is claims inflation in excess of normal economic inflation, over the period early 2006 to early 2008, and quite separate from the effect addressed in the preceding paragraph. The state space threshold model addresses this data pattern by allowing a flexible development year parameter (β_{ij} rather than β_j). Moreover, the threshold is estimated to be $T = 13$ which is at the beginning of 2006. Our model precisely detected this feature of the data. Adopting the same mean function throughout all accident quarters might fail to allow for the model shift after $T = 13$ caused by the lag effect of CLA 2003. Therefore, we introduce the state space threshold model to analyse this data. The adaptive nature of the state space threshold model will cause it to recognise the escalation of claims sizes eventually. However, there are others ways of constructing the model to recognise them immediately, such as a adding a time related terms in the model. This model will be explored in the future.

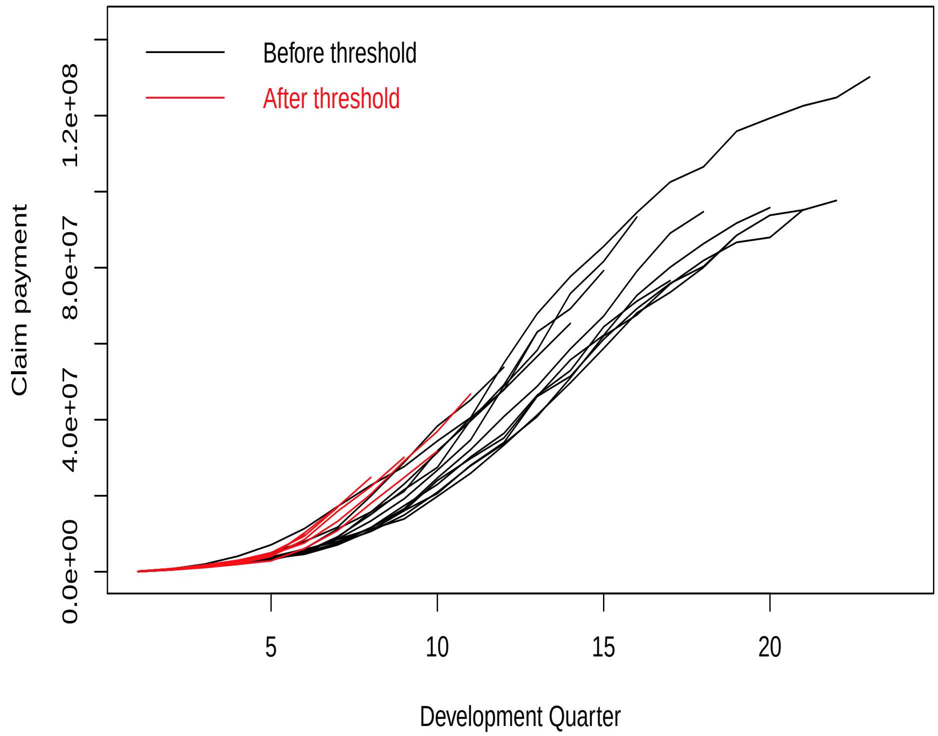


FIGURE 2.5. QLD CTP cumulative claims payment by development quarter

The threshold model can be used in conjunction with different kinds of mean models including the ANOVA (Model 2) and state space (Model 5) models. The state space threshold model is expressed as below:

$$
\begin{aligned}
Y_{ij} &\sim GB2(a, b_{ij}, p, q), \\
(a,p,q) &= \begin{cases} (a_1, p_1, q_1) & \text{for } i \le T, \\ (a_2, p_2, q_2) & \text{for } i > T, \end{cases} \\
\mu_{ij} &= \begin{cases} \mu_{1ij} = \upsilon_{10} + \alpha_{1i} + \beta_{1ij} & \text{for } i \le T, \\ \mu_{2ij} = \upsilon_{20} + \alpha_{2,i-12} + \beta_{2,i-12,j} & \text{for } i > T, \end{cases} \\
\alpha_{ki} &= \alpha_{k,i-1} + \upsilon_{ki}, \\
\beta_{kij} &= \beta_{k,i-1,j} + \nu_{ki}
\end{aligned}
\tag{16}
$$

where b_{ij} and $E(Y_{ij})$ are given by (6) and (7), $\upsilon_{ki} \sim N(0, \sigma^2_{\upsilon_k})$, $\nu_{ki} \sim N(0, \sigma^2_{\nu_k})$ and $\alpha_{k,i-1}$ and $\beta_{k,i-1,j}$ satisfy the constraints:

$$\alpha_{11} = \beta_{11,j} = \alpha_{21} = \beta_{21,j} = 0.$$

Alternatively, simplified ANOVA threshold (Model 3) and state space threshold (Model 6) models could be considered by setting the shape parameters of the GB2 distribution to be consistent across the threshold ($T = 13$); that is,

$$a_1 = a_2, \qquad p_1 = p_2, \qquad q_1 = q_2, \tag{17}$$

if they are similar across T for Models 2 and/or 5. Note that an alternative hierarchical form for Models 1 to 6 using the scales mixtures representation in (4) is

$$\begin{aligned} Y_{ij} &\sim GG(p, \lambda_{ij}, a), \\ \lambda_{ij} &\sim GG(q, b_{ij}, a) \end{aligned} \tag{18}$$

where b_{ij} and $E(Y_{ij})$ are given by (6) and (7), μ_{ij} for the ANOVA, state space and threshold state space models are given by (28), (32) and (33) respectively.

2.4.3. Numerical result. We start with fitting the traditional ANOVA model with three choices of distributions, Gamma, GT and GB2 distributions to the data. We have also used the chain ladder method, which is a widely recognized method of loss reserving for benchmarking.

The former two distributions are chosen because the Gamma distribution has been widely used by actuaries and the GT family does not nest within the GB2 family. Moreover models with the GB2 distribution and different mean functions are also attempted. Parameter estimates for Models 1 to 6 are reported in Table 1. Note that b_{ij} in (6) varies across accident quarter i and development quarter j for all models. The reported b is an average of b_{ij} over all quarters for Models 1 and 4. For the remaining models, the two b reported in Table 2.4.3 are averaged over accident quarters $T < 13$ and $T \geq 13$ respectively.

TABLE 2.1. Estimated GB2 Distribution Parameters

Models	a	b^*	p	q
M1 GB2 ANOVA	-8.67	19.3	0.95	21.80
M2 GB2 T ($i < 13$)	-9.073	31.0	1.55	3.24
($i \geq 13$)	-7.59	6.0	1.49	12.42
M3 GB2 T ($i < 13$)	-9.08	31.1	1.54	2.61
($i \geq 13$)	-	6.8	-	-
M4 GB2 SS	-8.68	17.9	1.17	39.38
M5 GB2 SS T ($i < 13$)	-4.95	24.1	1.95	9.12
($i \geq 13$)	-4.80	5.9	1.84	7.42
M6 GB2 SS T ($i < 13$)	-4.96	24.1	1.97	9.20
($i \geq 13$)	-	5.2	-	-

$*$ in unit of millions.

From Table 2.4.3, the estimated values of a are always negative, which indicates inverse distributions; the estimate of p from Model 1 is close to 1, which implies the Burr type 12 error distribution. The GB2 distribution is the most suitable distribution for the remaining models. In other words, any conventional distributions within the GB2 family are less suitable and hence yield less accurate prediction in loss reserving.

To evaluate these models, two criteria, model-fit and prediction accuracy, are considered. To assess the model-fit, two criteria: the R percentage and deviance information criteria (DIC) are adopted. The R percentage is the mean of predicted over actual loss less one, which is a popular measure to quantify the difference between actual and predicted values whereas DIC originated by Spiegelhalter *et al.* (2002) is a Bayesian analogue of Akaike's Information Criterion (AIC) which is commonly used in Bayesian analysis. DIC consists of a measure of model fit which is the posterior mean deviance, and a measure of model complexity which is an estimate of the effective number of parameters. It has a competitive advantage over the traditional AIC as it is not only limited to nested models. The DIC is given by

$$DIC = -\frac{4}{M}\sum_{m=1}^{M}\sum_{i=1}^{23}\sum_{j=1}^{24-i}\ln\left[f(y_{ij}|\boldsymbol{\theta}^{(m)})\right] + 2\sum_{i=1}^{23}\sum_{j=1}^{24-i}\ln\left[f(y_{ij}|\overline{\boldsymbol{\theta}})\right] \tag{19}$$

where $\boldsymbol{\theta}^{(m)}$ denotes the vector of parameter estimates in the m-th iteration of the posterior sample $M = 1000$, $\bar{\boldsymbol{\theta}}$ denotes the posterior mean of $\boldsymbol{\theta}^{(m)}$ and $f(y_{ij}|\boldsymbol{\theta})$ represents the observed likelihood in (1) for each observation where b is given by (6) and μ_{ij} in $E(Y_{ij})$ is given by (28), (32) and (33) for the three types of mean models.

Predictive performance is assessed by comparing the predicted aggregated loss with the actual loss in the last diagonal of the triangle ($i = 1, \ldots, 23$ and $i+j = 24$). The predicted aggregated loss $\hat{y}_{ij} = E(y_{ij})$ are calculated using (7) where μ_{ij} is given by (28), (32) and (33) for the three types of mean models. To project the cumulative loss in the lower triangle, the parameter estimates $\beta_{2,i-12,j}$ where both accident and development quarters are beyond the threshold $T = 13$ are unavailable and they are estimated by $\beta_{1j}\,\frac{\beta_{2,j-1}}{\beta_{1,j-1}}$ for the threshold ANOVA model and $\beta_{1,i-12,j}$ for the threshold state space model. Then the total of aggregate loss (as at June 2008), the observed y_{ij} and predicted $\hat{y}_{ij}$ are calculated as

$$y_o = \sum_{i=1,j=24-i}^{23} y_{ij} \quad \text{and} \quad y_p = \sum_{i=1,j=24-i}^{23} \hat{y}_{ij}$$

and the ratio defined as $R = \dfrac{y_p}{y_o} - 1$ is reported in Table 2.4.3 together with y_p and DIC.

Model with the smallest DIC and R percentage is preferred. R percentage measures the prediction accuracy. The GB2 state space threshold model (M6) with the lowest R percentage demonstrates the best predictive power as shown in Table 2.4.3. The model fit is quantified by DIC. The GB2 state space model (M5) provides the best model fit with the lowest DIC. It is not surprising that Model 0a and 0b perform less favorably according to DIC because the GT family requires transformation of the data and Gamma is a special case of the GB2 family. From a distribution perspective, the GB2 (M2) model outperforms the Gamma (M0a) and GT (M0b) models given

TABLE 2.2. Model selection for aggregated data

Models	DIC	R (%)
M0a GT ANOVA	16,142	-4.58
M0b Gamma ANOVA	8,586	-7.10
M0c Chain Ladder	-	-1.87
M1 GB2 T	8,521	-2.98
M2 GB2 ANOVA	8,566	-2.04
M3 GB2 SS T	8,651	-0.58
M4 GB2 T †	8,504	-2.55
M5 GB2 SS	8,093	-2.48
M6 GB2 SS T †	8,566	0.54

† represents models with different p and q values before and after T.

the same ANOVA mean function; from a mean function perspective, the threshold (M4) and state space models (M5) outperform the ANOVA (M2) given the same GB2 distribution.

In Figure 2.6, we apply Model 6 to project the claims payment flight path. The predicted cumulative loss $\hat{y}_{ij}$ in the lower triangle are plotted in dotted black lines when $i < T$ and dotted red lines when $i \geq T$) whereas the observed y_{ij} in the upper triangle are plotted in solid lines. The figure demonstrates a general upward trend at a fast rate till the 15-th development quarter, a slow rate thereafter, and gradually level off. It also shows a distinct pattern before and after the threshold T: the later pattern shows a sharper upward trend at earlier development quarter.

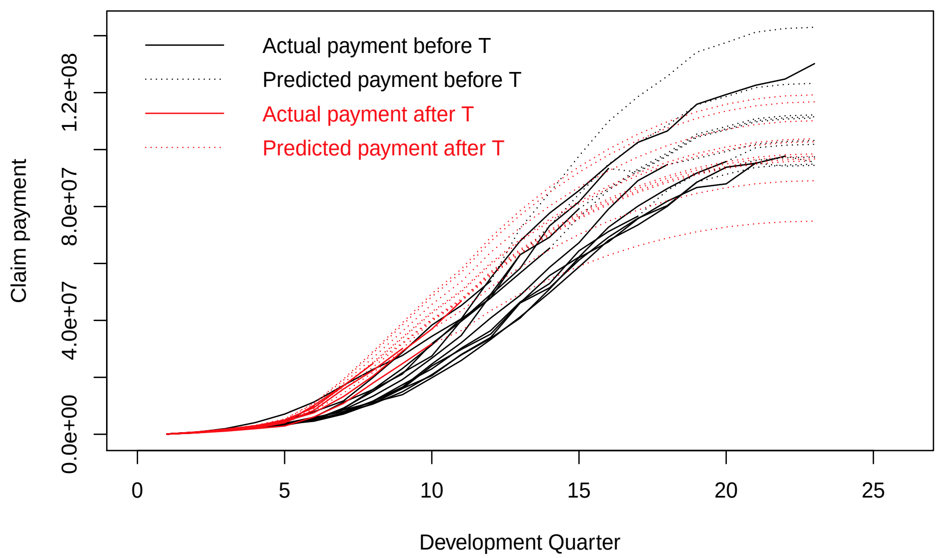


FIGURE 2.6. QLD Claims Payment Projection by Development Quarter

Figure 2.4.3 presents three triangular heat maps to visualize the ratio of fitted to actual loss by accident and development quarters in the upper triangle for three models. Generally speaking, green represents good fit; yellow color indicates overprediction whereas blue color reveals underprediction. There is more green cells in the first graph than the second one, which implies that the GB2 ANOVA model provides better fit than the Gamma ANOVA model given the same ANOVA mean function; with more blue cells in the second graph, the Gamma model shows predominately underprediction. Comparing the first and third graphs, the GB2 state space threshold model (M6) with more green cells clearly demonstrates the best fit to the data, which is in conjunction with the results in Table 2.4.3. Some common patterns are observed in the three heat maps. For example, the underprediction in the first 4 to 8 development quarters of the accident quarter 11 (Jun-05) exists in all three graphs, and also the underprediction from the 6 to 8 development quarters of accident quarter 1 (Dec-02) in the first and second graphs, but became much less in the third graph. One possible reason is that the Civil Liability Act 2003 applied to accident dates on and after the first of December 2002. So the majority of accident quarter 1 is subject to a legislative regime different from that applying to all subsequent accident quarters. The state space threshold model represented by the third graph offers different parameters for each development and accident quarters, thereby

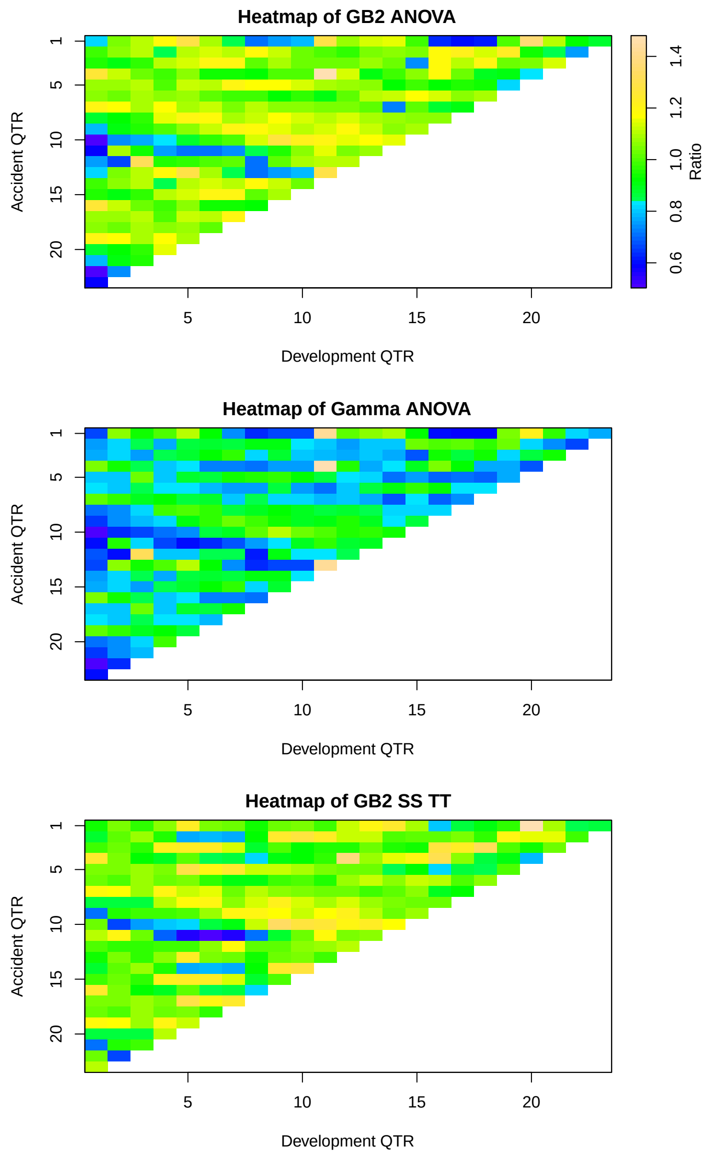


FIGURE 2.7. Triangular Actual vs Expected Heat map.

recognising these patterns well. Furthermore, the threshold models assume a threshold at accident quarter 13 (Dec-05), which coincides with a number of significant changes in claims experience occurred around this period. It is generally agreed by actuaries for Queensland CTP insurers and for the Queensland regulator that the relative incidence of lower severity claims increased, rates of claim finalisation changed in response, and that claim sizes were affected by very high super-imposed inflation between payment quarters Jun-06 to Mar-08. The threshold models, to some extent, recognise these changes.

TABLE 2.3. Simulation Results

Parameters	a	p	q	μ_0	α_1	β_1
True values	-8.667	0.946	21.800	2.009	0.315	-5.895
Estimated values	-9.105	1.108	23.017	2.067	0.317	-5.892
APB	0.051	0.171	0.056	0.029	0.005	0.000
SD	1.507	0.371	4.485	0.019	0.031	0.034

2.4.4. Simulation study. In this simulation study, we evaluate the performance of the basic GB2 ANOVA model. The performance of other GB2 models will be similar due to the same distributional assumption. The GB2 ANOVA model has 48 parameters. We use the estimated parameter values to simulate $N = 200$ data sets; each contains $n = 276$ observations, which is the same as the size of the QLD CTP payment data. Table 2.4.4 reports the mean, absolute percentage bias (APB), and standard deviation (SD) of the parameter estimates over $N = 200$ replications. APB is expressed as the absolute value of the difference between estimated and true value as a proportion of the true value. The results show that the parameters involved in the mean function including μ, α_i and β_j achieve high level of accuracy whereas the shape parameters are estimated to a moderate level of accuracy. It is well known that shape parameters are often more difficult to estimate because distributions are less sensitive to some shape parameters. However, as the main focus of the analysis is on the projection of loss reserve based on the expected mean value of each cell in the loss triangle, the shape parameter estimates have minimal effect on the projection. In summary, the model performance is satisfactory, and therefore the parameter estimates and forecasts in the empirical study of aggregated data are reliable.

2.5. Study of individual loss data

Model for individual loss data has been increasingly adopted to analyze the effect of individual characteristics because it helps to identify the underlying drivers of claim cost. It provides not only a link between changes in the claims processes and reserving, but also an understanding and quantification of the drivers of a claim. Moreover, loss reserve models for individual claims provides individual estimates of future claim costs arising from existing claims. These individual predictions form the basis of reinsurance premium calculations; it is as important as total reserve estimation.

2.5.1. The data. In this study, the data that we analyze refers to the workers' compensation (WC) journey claims which have been re-directed to CTP insurers in QLD as at June 2008. It consists of 2516 individual claims. In order to predict the full claim cost, the data include only the finalized claims and are inflated to December 2008. Three key variables are selected for the WC claims predictors. Their definitions and levels are given as follows:

Finalization delay (F_i): The number of months taken for the claims to be finalized

Role of claimants: The role of claimants in an accident (Levels: Driver (Dri; $x_{1i} = 1$, $x_{2i} = 0$), motercycle rider (Mcr; $x_{1i} = 0$, $x_{2i} = 1$), (Passenger; $x_{1i} = x_{2i} = 0$) For each claim i, exactly one of the x_{ki} takes a nit value and the remainder are equal to zero.

Treatment Indicator: The length of treatment the claimant received (Levels: Short term ($\leq$ 6 weeks; $x_{3i} = 1$), long term (>6 weeks; $x_{3i} = 0$))

The Ninety, Ninety-five and ninety-nine percentiles of the data are 13,657, 46,424 and 325,322 dollars respectively, showing a dramatic increase from ninety-five to ninety-nine percentiles. A preliminary data analysis reveals that the maximum value of the data is 85 times the mean. Moreover, the mean, standard error, skewness and kurtosis for the data are 3839, 13955, 14.09 and 258.38, respectively and the claims above 95 percentile account for 50.4 percent of the total claims cost. All these features present strong evidence that the WC data are heavily tailed and exhibit considerable heterogeneities. Modelling the tail behavior is sure to have high impact on the accuracy of loss reserve.

2.5.2. The mixture model. In general insurance practice, small and large CTP claims present distinct risk characteristics based on managers' claims handling experience. Consequently, CTP claims managers usually handle small and large claims separately. We adopt an alternative modelling approach in which a single model is applied to the whole data set with the characteristics of both groups. From a modelling perspective, this allows share of information, resulting in a more efficient model.

To allow for such heterogeneity in the WC data, we assume there are two underlying subgroups in the mixture of GB2 model. It captures group-specific characteristics arising from the low and high level loss payments in the WC data. Besides, the model facilitates classification of loss payment into different risk groups, thereby providing insurance companies greater insight so as to distinguish claims at an earlier stage.

Suppose that there are two risk groups and each claim has a probability, $\pi_k \geq 0$ of coming from group $k, k = 1, 2$ and $\pi_2 = 1 - \pi_1$. We define the unobserved group-k membership indicator $I_{ki} = 1$ if a claim i belongs to group k and $I_{ki} = 0$ otherwise and it follows Bernoulli distribution with probability π_k. If a claim Y_i belongs to group k,

$$
\begin{aligned}
Y_i &\sim GB2(a_k, b_{ik}, p_k, q_k), && (20)\\
b_{ik} &= \frac{E(Y_i)B(p_k, q_k)}{B(p_k + 1/a_k, q_k - 1/a_k)}, && (21)\\
log(E(Y_i)) &= \mu_k + \alpha_k F_i + \beta_{1k}x_{1i} + \beta_{2k}x_{2i} + \beta_{3k}x_{3i}. && (22)
\end{aligned}
$$

Then Y_i is said to arise from a finite mixture model with probability density function $f(y)$:

$$f(y) = \pi_1 * f_1(y) + (1 - \pi_1) * f_2(y),$$

where $f_k(y)$ is the probability density function (1) for component k with parameters $\boldsymbol{\theta}_k = (a_k, p_k, q_k, \mu_k, \alpha_k, \beta_{1k}$ and together with the missing observation I_{ki}, the whole vector of parameters $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \pi_1, \boldsymbol{I}_1)$ where $\boldsymbol{I}_1 = (I_{11}, \ldots, I_{1n})$, π_1 is the weight. Note that $\boldsymbol{\Theta}$ contains two types of parameters, the model parameters such as p_k and β_{jk} and the missing group membership I_{1i}. Both of them are estimated by drawing samples from their conditional posterior distributions.

To derive posterior distributions for the model parameters, the following diffuse priors are assigned to the above modelling parameters. The weight π_k is assigned a non-informative Uniform prior on the range from 0 to 1, as it represents the weights between the two components of the mixture model. The shape parameters p_k and q_k are assigned Gamma distributions with different sets of hyper-parameters, to allow for more flexible tail behavior which is described by the ratio between p_k and q_k. The mean and variance of q_k are set to be larger than that of p_k as q_k is more sensitive and tends to take greater values from experience.

$$
\begin{gathered}
\mu_k \sim N(0, 100), \quad \alpha_k \sim N(0, 100), \quad \beta_{jk} \sim N(0, 100), \quad \pi_k \sim U(0, 1), \\[1em]
a_k \sim N(0, 100), \qquad p_k \sim Ga(0.001, 0.001), \qquad q_k \sim Ga(0.01, 0.0001)
\end{gathered}
$$

where $U(a, b)$ denotes the uniform distribution with support a to b.

One issue we encountered in implementing the mixture of GB2 distribution via the Markov chain Monte Carlo (MCMC) Bayesian approach is the label switching problem. It is primarily caused by the likelihood of a mixture model being invariant to permutations of the labels. The permutation can change many times across MCMC iterations making it difficult to infer component-specific parameters of the model. Lee *et al.* (2008) solved the label switching problem by imposing identifiability constraints on the parameters in a normal mixture model. We adopt the same idea by constraining the intercept of the first group to be smaller than that of the second group; that is $\mu_1 < \mu_2$. In other words, μ_2 is sampled from the range of (μ_1, ∞). This constraint ensures the vector of parameter estimates corresponds to its unique claims group in each MCMC iteration. The adoption of this identifiability constraint has substantially stabilized the parameter estimates in the simulation process resulting in more reliable measures of component specific effects.

2.5.3. Numerical result. We start with Gamma and GB2 error distributions without mixture effects. We then consider the proposed mixture model. The parameter estimates in the mean function for all three models are reported in Table 2.5.3. The direction of effects in the mean function across models is consistent for the four variables. Longer finalization delay and length of treatment lead to higher claim cost; the costs associated with drivers are less than those of passengers which in turn is less than those of motorcycle riders. Classification of loss payments into the two distinct risk groups can be performed using I_{ik}: loss payment i is assigned to risk group k if $\bar{I}_{ik} > 0.5$. For the mixture GB2 model, the first group consists of lower claims with a mean of 1621.5 and accounting for 60% of claims whereas the second group contains mainly large claims with a mean of 8405.0. From checking the credible intervals for the differences in parameter estimates across the two groups, we found that α and β_3 are significantly different.

To compare these models, we evaluate the model-fit using DIC. Celeux et al. (2002) advanced the DIC for mixture model as follows:

$$
\begin{aligned}
DIC \;&=\; -\frac{4}{M}\sum_{m=1}^{M}\sum_{i=1}^{n}\sum_{k=1}^{G} I_{ki}^{(m)} \ln\left[\pi_k^{(m)} f(y_i|\boldsymbol{\theta}_k^{(m)})\right] + \\
&\quad\; 2\sum_{i=1}^{n}\sum_{k=1}^{G} \bar{I}_{ik} \ln\left[\overline{\pi}_k f(y_i|\overline{\boldsymbol{\theta}}_k)\right] \qquad (23)
\end{aligned}
$$

where $\boldsymbol{\theta}_k^{(m)}$ denotes the vector of model parameter estimates in the m-th iteration of the posterior sample for group k, G is the number of groups, $I_{ik}^{(m)}$ denotes the estimate of I_{ik} in the m-th iteration, $\bar{\boldsymbol{\theta}}_k$ and $\bar{I}_{ik}$ denotes the posterior mean of $\boldsymbol{\theta}_k^{(m)}$ and $I_{ik}^{(m)}$ respectively. $f(y_i|\boldsymbol{\theta}_k)$ represents the observed likelihood in group k where $E(Y_i)$ in b_{ik} is given by (22). Parameter estimates for the three models and measures of model-fit including RMSE and DIC are listed in Table 2.5.3. The results show that DIC favors the GB2 model whereas root mean square error $RMSE$ indicates the mixture GB2 model provides the best fit. The mixture GB2 model provides less satisfactory DIC because all group membership indicators I_{1i} are considered as parameters, thereby substantially inflating the number of parameters for the mixture GB2 model as compared to the GB2 model without mixture effects. The model fit component of DIC for the mixture GB2 model is -46,781, which is in fact less than -44,143 for the GB2 model (G=1). Moreover, Figure 2.8 clearly shows that the mixture GB2 model outperforms the model using a single GB2 distributions in modeling the peak for the small claims and the tail for the large claims.

TABLE 2.4. Parameter estimates for the mean function

Models	Intercept μ	Fin. del. α	Role: Dri. β_1	Role: Mcr β_2	Tre: Sho. β_3	Prob. π
Gamma	7.520	0.040	-0.404	0.587	-0.145	-
GB2	9.273	0.015	-0.162	0.198	-0.131	-
Mixture GB2 (lower)	7.342	0.004	-0.089	0.036	-0.034	0.592
Mixture GB2 (higher)	8.165	0.045	-0.357	0.309	-0.425	0.408

TABLE 2.5. Parameter estimates and model-fit measures for individual loss data

Models	a	b	p	q	$RMSE$	DIC
Gamma	6.99†	2.00	-	-	16,454	47,832
GB2	-2.04	502.26	0.52	1.72	16,351	44,152
Mixture GB2 (lower)	-2.67	127.97	0.56	44.22	15,634	56,367
Mixture GB2 (higher)	-1.29	4,594.94	1.35	1.16	-	-

In the Gamma model, a and b denote the scale and shape parameters respectively;

† the value of a should be multiplied by 10,000.

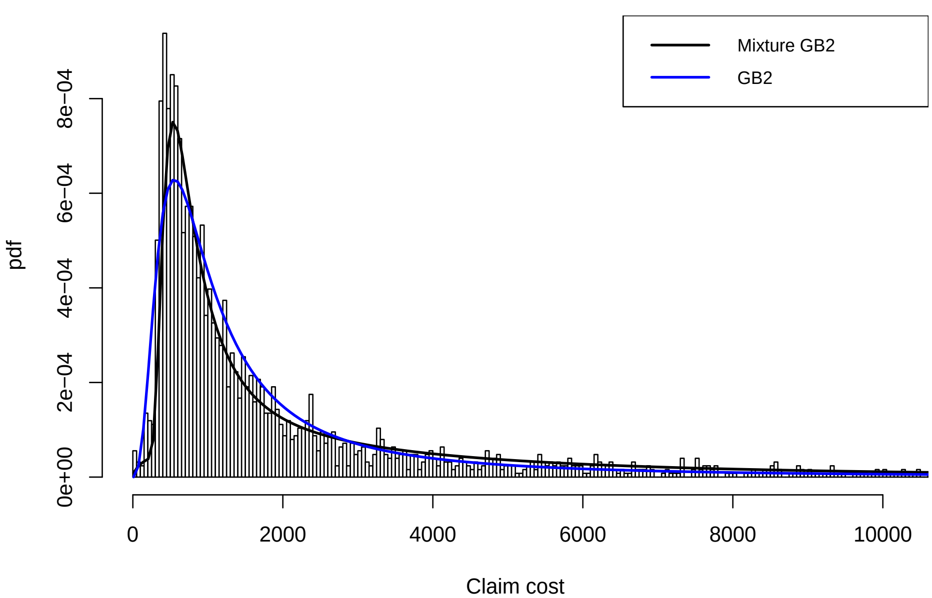


FIGURE 2.8. Density of mixture GB2 and GB2 models.

In the mixture GB2 model, the parameter estimate of q in the lower level group with a large value reveals that the distribution approaches generalized gamma whereas that of the higher level group (q close to 1) tends to Burr type 3. In all cases, a possesses negative values which imply inverse distributions. Note that the values of b are averages over all observations for the first two models in Table 2.5.3. In the mixture case, the two b are further weighted by the estimates of the group membership indicators $\bar{I}_{ki}$. Estimates of all shape parameters ($a,\ p$ and q) are found to be significantly different between risk groups as revealed by the credible intervals. Since most of the parameter estimates are significantly different across the two groups, there is a strong evidence that the two group mixture model is most suitable for the WC data. Figure 2.9 provides a graphical illustration that the density curve of the mixture of GB2 distributions closely envelops the empirical data histogram. The lower level group (in blue) accounts for the peak of the density whereas the higher level group (in red) has a thicker tail than that of the lower level group to accommodate extremely large claims. The two groups compensate the deficiency of each other in the combined mixture model (in black).

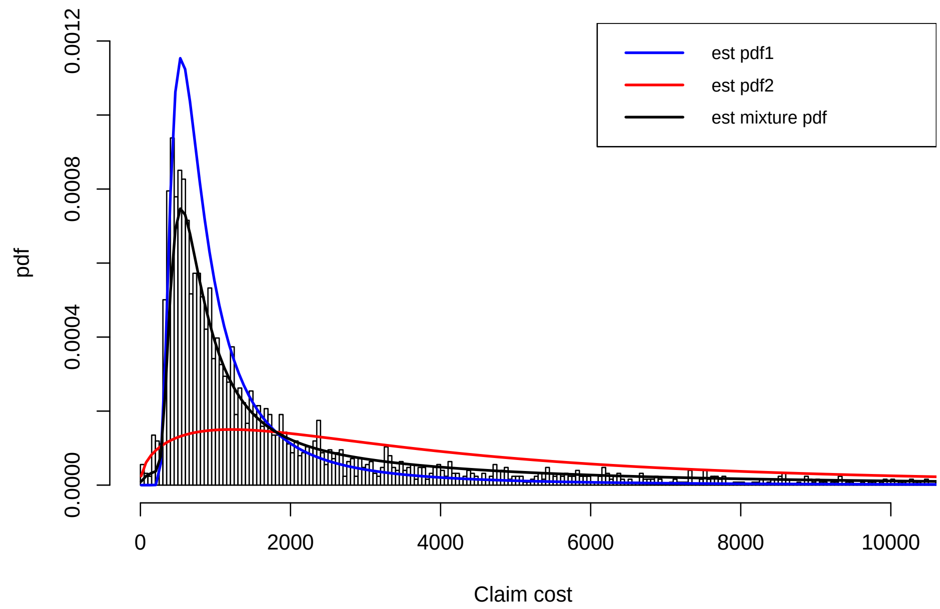


FIGURE 2.9. Density of mixture GB2 distribution and its components.

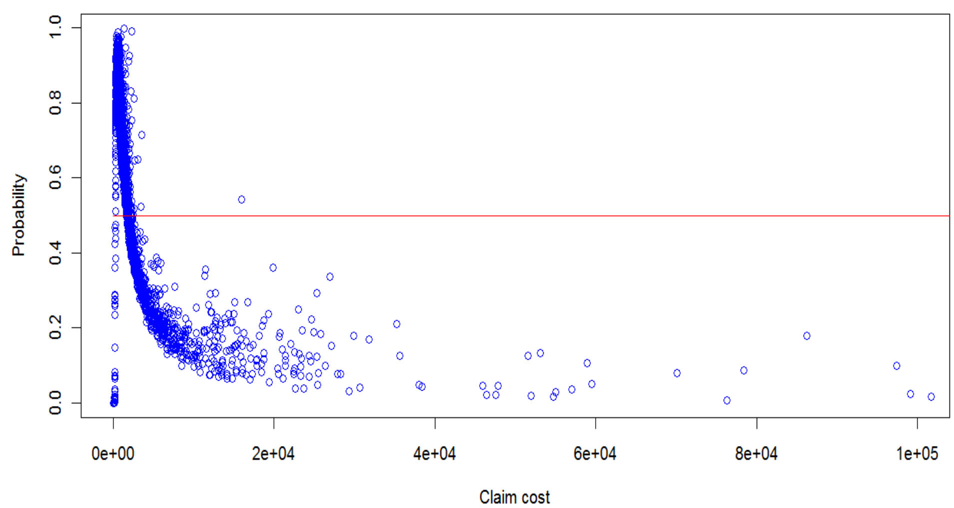


FIGURE 2.10. Membership of the lower level claims group.

As mixture model enables identification of claims groups using $\bar{I}_{ki}$, it provides insight for insurance company to manage claims. Figure 2.10 plots the indicator estimates $\bar{I}_{1i}$ for the lower level claim group. The red line indicates the cut off between lower and higher level claims groups. We can see that as claims cost increases, the proportion of claims classified to the lower claims group decreases.

2.6. Conclusions

In this chapter, we explore dynamic modeling for the long tail loss reserving data. The flexible modelling approach presented in section 4 allows for the distributional parameters and means to evolve with respect the actual circumstances of the data. We demonstrate the threshold and state space mean models under the GB2 distribution tailoring to the CTP loss reserving data under the influence of legislation changes. The GB2 distribution is shown to encompass many useful heavy and light tail distributions, and hence provides sufficient flexibility to address the tail where a substantial amount of losses are above the 95 percentile. Although Cummins et al. (2007) applied GB2 distribution to the claims data, they fitted separate distributions to the claims in each cell of the payout triangle to allow for the change in risk over time and across lags. The parameters estimated using this approach are only based on the data in one cell, and hence resulting in less reliable estimates because of the limited data size and ignorance of dependency between cells. Our approach significantly improves their model by utilizing all the data in the payout triangle, and at the same time allowing for any unexpected change across accident or development years. The simulation study in section 2.4.4 confirms that our parameter estimates are very reliable and achieve high level of accuracy.

The second part of this chapter (section 2.5) presents the GB2 distribution under the mixture framework for the individual loss data tailoring to its heterogeneous features. The mixture framework allows for any parameters to vary across different risk groups. In real insurance practice, mixture model has major advantage over traditional models as it models unobserved heterogeneity in the data and enables classification of claims into different risk groups. The resulting groups provide insight for insurance companies for better claim management. On the other hand, it allows actuaries to focus on the high risk group for accurate projection of large claims cost, which has substantial impact on expected severities under reinsurance contracts and recoveries calculations. To facilitate the implementation of the proposed models, we present the Bayesian hierarchy using WinBUGS. Moreover, we introduce the alternative hierarchical forms of the models using the scales mixtures representation for the GB2 distribution to simplify the Gibbs sampler in the MCMC simulation.

Although our proposed model is flexible enough to model claims in many real situations, there are certainly ways for further improvements. In particular, we can explore the possibilities of allowing the shape of the distribution to change across each risk combination by adopting more flexible forms of the distributional parameters. It can be achieved by equating functions of individual claim characteristics, accident and development quarters etc., not only to the mean, but also to the distributional parameters. This is equivalent to building one model for each risk combination but has the benefit of incorporating all data. Yang et al. (2011) consider a multivariate GB2 model to capture non-elliptical and asymmetric dependencies among claim portfolios. Extending our proposed approaches to multivariate modelling of claims will surely increase the applicability of the models considerably. Although the full extent of the dynamic models is still to be evaluated, our results do show promise.

CHAPTER 3

Risk Margin Quantile Regression Model

Continued from the development of mean models to quantile functions to derive risk margin and to evaluate capital in non-life insurance applications via parametric and nonparametric quantile regression models.

3.1. Background on Risk Margin Calculation

A core component of the work performed by general insurance actuaries involves the assessment, analysis and evaluation of the uncertainty involved in the claim process with a view to assessing appropriate risk margins for inclusion in insurance liabilities. An appropriate valuation of insurance liabilities including risk margin is one of the most important issues for a general insurer. Risk margin is the component of the value of claims liability that relates to the inherent uncertainty.

The significance of this task is well understood by the actuarial profession and has been debated by both practitioners and academic actuaries alike. Much of the attention involves the non prescriptive nature of risk margin requirements discussed in regulatory guidelines such as Article 77 and Article 101 of the Solvency II Directives. In Australia a general task force was established, developing a report on risk margin evaluation methodologies presented to the Australian actuarial profession at the Institute of Actuaries of Australia during the 16-th General Insurance Seminar in 2008. This report aimed to highlight approaches to risk margin calculations that are often considered. Before briefly discussing these aspects we first note the following Solvency II items which relate to the Solvency Capital Requirement (SCR) and the risk margin.

Article 101 of the Solvency II Directive states,

> “ *The Solvency Capital Requirement (SCR) shall correspond to the Value-at-Risk (VaR) of the basic own funds of an insurance or reinsurance undertaking subject to a confidence level of 99.5% over a one-year period.* ”

Essentially, the basic own funds are defined as the excess of assets over liabilities, under specific valuation rules. In this regard, a core challenge is the capital market-consistent value of insurance liabilities, which requires a best estimate typically defined as the expected present value of future cash flows under Solvency II plus a risk margin calculated using a cost-of-capital approach.

Furthermore, under Article 77 of the 2009 Solvency II Directive it states that the risk margin calculation is described as

" *The risk margin shall be such as to ensure that the value of the technical provisions is equivalent to the amount insurance undertakings would be expected to require in order to take over and meet the insurance obligations... âĂę it shall be calculated by determining the cost of providing an amount of eligible own funds equal to the Solvency Capital Requirement necessary to support the insurance obligations over the lifetime thereof...* "

As can be seen from such specifications, the recommendations to be adopted are not prescriptive in the required model approaches. Therefore, as discussed in the white paper produced by the Risk Margins Task force 1998, there have been several approaches considered which range from those that involve little analysis of the underlying claim portfolio to those that involve significant analysis of the uncertainty using a wide range of information and techniques, including stochastic modelling. They highlighted approaches adopted in practice in the assessment of risk margins and pointed to percentile or quantile methods as being most prevalent in practice, this provides a good foundation for the methods we consider.

Traditionally, actuaries that adopt a stochastic framework would evaluate claims liability using a central estimate which is typically defined as the expected value over the entire range of outcomes. However with the inherent uncertainty that may arise from such an estimator which is not statistically robust and therefore sensitive to outlier claims, claims liability measures often differ from their central estimates. In practice, the approach adopted is typically to then set an insurance provision so that, to a specified probability, the provision will eventually be sufficient to cover the run-off claims. For instance, in order to satisfy the requirement of the Australian Prudential Regulation Authority (APRA) to provide sufficient provision at a 75% probability level, the risk margin should be modelled statistically so that it can capture the inherent uncertainty of the mean estimate. When this margin is then added to the central estimate, it should provide a reasonable valuation of claims liability and therefore increases the likelihood of providing sufficient provision to meet the level required in GPS 320. In this regard, it is worth noting that the more volatile a portfolios runoffs or those that display heavy tailed features may require a higher risk margin, since the potential for large swings in reserves is greater than that of a more stable portfolio.

To accommodate these ideas, two common methods for risk margin estimation have been proposed in practice. These are the cost of capital and the percentile methods. Under the cost of capital method the actuary determines the risk margin by measuring the return on the capital required to protect against adverse development of those unpaid claim liabilities. It is evident that application of the cost of capital method requires an estimate of the initial capital to support the unpaid claim liabilities and also the estimate of return on that capital. Alternatively, under the percentile or quantile method that we consider in this chapter, which is currently used in Australia the actuary takes the perspective that the insurer must be able to meet its liability with some probability under some assumptions on the distribution of liabilities. Risk margin is then calculated by subtracting the central estimate from a predefined critical percentile value.

What we bring to the percentile and quantile based framework in our proposed methods is the ability to incorporate in a rigorous statistical manner, regression factors that may be related to both exogenous features directly related to the insurance claims run-off stochastic process as well as endogenous factors that are related to for instance the current micro or macro economic conditions and the regulatory environment. These will be incorporated into a statistical model that allows one to explain the proportion of variation in the risk margin that is attributed to such features in a principled manner, as we shall demonstrate allowing for accurate estimation and prediction. We argue that since the percentile-based method involves the estimation of quantiles, it is therefore somewhat natural to consider quantile regression, which is a statistical technique to estimate conditional quantile functions, which can be used to estimate risk margin.

Just as classical linear regression methods based on minimizing sums of squared residuals enable one to estimate models for conditional mean functions, quantile regression methods offer a mechanism for estimating models for the conditional median function, and the full range of other conditional quantile functions. This model allows studying the effect of explanatory variables on the entire conditional distribution of the response variable and not only on its center. Hence we may develop factors and covariates which are explanatory of the risk margin variation directly through the proposed quantile regression framework. By supplementing the estimation of conditional mean functions with techniques for estimating an entire family of conditional quantile functions, quantile regression is capable of providing a more complete statistical analysis of the stochastic relationships among random variables.

Quantile regression has been applied to a wide range of applications in economics and finance, but has not yet been developed in a claim reserving context for risk margin estimation. We will demonstrate the features of quantile regression that have been popularized in finance and explain how they can be adopted in important applications in insurance, such as risk margin calculations. In quantitative investment, least square regression-based analysis is extensively used in analyzing factor performance, assessing the relative attractiveness of different firms, and monitoring the risks in their portfolios. Engle and Manganelli (2004) consider the quantile regression for the Value at Risk (VaR) model. They construct a conditional autoregressive value at risk model (CAVaR), and employ quantile regression for the estimation. The risk measure, VaR is defined as a quantile of the loss distribution of a portfolio within a given time period and a confidence level. Accurate VaR estimation can help financial institutions maintain appropriate capital levels to cover the risk from the corresponding portfolio.

Taylor (2006) estimate percentile-based risk margins via a parametric model based on the assumption of a log normal distribution of liability. Other sophisticated distributions to capture flexible shapes and tail behaviors are also proposed to model severity distribution on aggregated claim data. These distributions include the generalized-t (McDonald and Newey, 1988), Pareto (Embrechts et al., 1997), the Stable family (Paulson and Faris, 1985; Peters et al., 2011b,c), the Pearson family (Aiuppa, 1988), the log-gamma and lognormal (Ramlau-Hansen, 1988), and type II generalized beta (GB2) distribution (Cummins et al., 2007). While these distributions on real support are flexible to model both leptokurtic and platykurtic data, they require log-transformation for claims data and the resulting log-linear model may be more sensitive to low values than large values (Chan et al., 2008).

In Peters et al. (2009) they adopt a Poisson-Tweedie family of models which incorporates families such as normal, compound poisson Gamma, positive stable and extreme stable distributions into a family of models. It was shown how such a generalized regression structure could be used in a claims reserving setting to model the claims process whilst incorporating covariate structures from the loss reserving structure. In this instance a multiplicative structure for the mean and variance functions was considered and quantiles were derived from modelling the entire distribution, rather than specifically targeting a model at the conditional quantile function.

Recently, in Dong and Chan (2013) an alternative class of flexible skew and heavy tail models was considered involving the GB2 distribution with positive support adopting dynamic mean functions and mixture model representation to model long tail loss reserving data and showed that GB2 outperforms some conventional distributions such as Gamma and generalised Gamma. The GB2 distribution family is very flexible as it includes both heavy-tailed and light-tailed severity distributions, such as gamma, Weibull, Pareto, Burr12, lognormal and the Pearson family, hence providing convenient functional forms to model claims liability. From the perspective of quantile specific regression models, recently **?** proposed a power-Pareto model which allows for flexible quantile functions which can provide a combination of quantile functions for both power and Pareto distributions. These combinations enable the modelling of both the main body and tails of a distribution.

The difference with our current methodology is that instead of developing a statistical model to capture all features of the claims run-off stochastic structure, with the incorporation of regression components, we propose, in this work, to target explicitly the conditional quantile functions in a regression structure. From a statistical perspective, this is a fundamentally different approach to these previously mentioned reserving model approaches. However we will illustrate that we can borrow from such models in developing our risk margin quantile regression framework. In fact the associate parameter estimation loss functions, parameter estimator properties and the resulting quantile in sample and out of sample forecasts will significantly differ to those achieved when trying to develop a model for the entire process rather than targeting the quantity of interest in this case, the particular quantile level. This is clear from the perspective that only under a Gaussian distributional assumption for such reserve models (on log scale) would a standard least squares approach be optimal from the perspective of Gauss-Markov theory. In situations where returns are heavy tailed and skewed alternative models will prove more appropriate as we will discuss.

Traditional approaches, both frequentist and Bayesian, to quantile regression have involved parametric models based on the asymmetric Laplace (AL) distribution. Using asymmetric Laplace distribution provides a mechanism for Bayesian inference of quantile regression models. Hu et al. (2012) develop a fully Bayesian approach for fitting single-index models in conditional quantile regression. The benefit of using a Bayesian procedure, lies in the adoption of available prior information and the provision of a complete predictive distribution for the required reserves (De Alba, 2002). Different Bayesian loss reserve models have been proposed for different types of claims data. Zhang et al. (2012) propose a Bayesian non linear hierarchical model with growth curves to model the loss development process, using data from individual companies forming various cohorts of claims. Ntzoufras and Dellaportas (2002) investigate various models for outstanding claims problems using a Bayesian approach via Markov chain Monte Carlo (MCMC) sampling strategy and show that the computational flexibility of a Bayesian approach facilitated the implementation of complex models.

3.1.1. Contributions. The contribution of this chapter is three-fold. First, we propose using quantile regression for loss reserving. The proposed method, relating the provision to quantile regression allows a direct modelling of risk margin, and hence provision, instead of estimating the mean then applying a risk margin. It provides a richer characterization of the data, especially when the data is heavy tailed, allowing us to consider the impact of a covariate on the entire distribution, not merely its conditional mean. Secondly, we develop a range of parametric quantile regression models in Bayesian framework, each with their own distribution features. Especially, in particular we generalize the AL distribution model to incorporate a dynamic mean, variance and the shape parameters to model risk margin via a user friendly Bayeisan software `WinBugs`, which is easy for users without much Bayeisan background or specialized knowledge of Markov chain Monte Carlo (MCMC) methodology. Furthermore, the estimation of shape parameter by accident year gives us an analytical framework to estimate risk margin. This allows us to capture the feature that the cohort of claims in different accident year may be heterogeneous, and hence applying different different risk margin to different accident year gives us an explicit provision in reserving. Finally, we compare the performance of parametric and nonparametric quantile regressions in the context of loss reserving.

The rest of the chapter is organized as follows. Section 2 explains the parametric and nonparametric models proposed. Section 3 presents the posterior quantile regression models in a

Bayesian framework. Section 4 details the way to calculate risk measures and risk margin using our models. Then we apply the methodology to two real loss reserve data sets in Section 5 and 6. Section 7 concludes.

3.2. Quantile Regression for Claims Reserving

In this section, we present quantile regression models and explain their relevance to loss reserving, this will be undertaken in both a non-parametric and a parametric modelling framework under the Bayesian paradigm. In the process we propose a novel analytical approach to perform estimation of the risk margin under various quantile regression model structures. Of particular focus in this chapter is the class of models based on the Asymmetric Laplace (AL) distributional family. In the special case of the AL distribution we demonstrate that risk margin estimation is achieved naturally through the modelling the shape parameters of the AL distribution and hence the inference on the model parameters directly informs the inference of the risk margin.

In developing a quantile regression framework for general insurance claims development triangles we will assume that there is a run-off triangle containing claims development data in which Y_{ij} will denote the cumulative claims with indices $i \in \{0, ..., I\}$ and $j \in \{0, ..., J\}$, where i denotes the accident year and j denotes the development year (cumulative claims can refer to payments, claims incurred, etc). Furthermore, without loss of generality, we make the simplifying assumption that the number of accident years is equal to the number of observed development years, that is, $I = J$ with $N = \frac{1}{2}I(I+1)$ observations. At time I the index set in the *upper* triangular is

$$\mathcal{D}_o = \{(i,j):\ i + j \leq I + 1\} \tag{1}$$

and for claims reserving at time I the index set to predict the future claims in the *lower* triangle is:

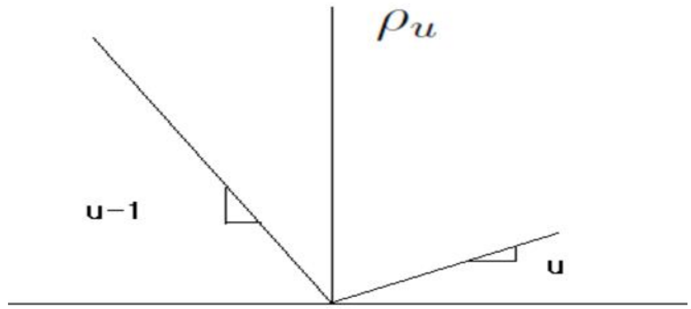
$$\mathcal{D}_l = \{(i,j):\ i + j > I + 1,\ i \leq I,\ j \leq I\}\,. \tag{2}$$

Therefore the vector of observed Y_{ij} in the upper triangle is given by $\boldsymbol{Y}_o = \{Y_{ij} : (i,j) \in \mathcal{D}_o\}$ and the corresponding vector of covariates is denoted by $\boldsymbol{x}_o = \{\boldsymbol{x}_{ij} : (i,j) \in \mathcal{D}_o\}$. Similarly $\boldsymbol{Y}_l = \{Y_{ij} : (i,j) \in \mathcal{D}_l\}$ and $\boldsymbol{x}_l = \{\boldsymbol{x}_{ij} : (i,j) \in \mathcal{D}_l\}$ are the vectors of claims and covariates in the lower triangle.

In the quantile regression structures we will aim to make inference on the quantile function of the data within sample, in each cell of $\boldsymbol{Y}_o$ as well as predictive out-off sample quantile function estimation based on the claim cells in $\boldsymbol{Y}_l$ in lower triangle. The estimation of the quantile function regression has three main components:

- The conditional distribution and in this case conditional quantile function of the dependent variables given by the claims data, given the explanatory variables.;

FIGURE 3.1. Loss function



- The structural component of the regression structure based on the link functions and imposed model structures linking the regression structures with the covariates to the location and scale of the conditional distribution and conditional quantile functions of the response.;
- The actual choice of independent variables i.e. the covariates in the regression model, in this case we will also consider some basis function regression structures in some of the models proposed.

In the following sub-sections we discuss each of these components in term, starting with the distributional aspects of the quantile regression models we consider.

3.2.1. Nonparametric Quantile Regression Models. In a non-parametric quantile regression approach, we perform estimation of regression coefficients without the need to make any assumptions on the distribution of the response, or equivalently the residuals. If $Y_{ij} > 0$ is a set of observed losses and $\boldsymbol{x}_{ij} = (1, x_{ij1}, \ldots, x_{ijm})$ is a vector of covariates that describe Y_{ij}. The quantile function for the log transformed data $Y^*_{ij} = \ln Y_{ij} \in \Re$ is

$$Q_{Y^*}(u|\boldsymbol{x}_{ij}) = \alpha_{0,u} + \sum_{k=1}^{m} \alpha_{k,u}\, x_{ijk} \tag{3}$$

where $u \in (0,1)$ is the quantile level, $\boldsymbol{\alpha}_u = (\alpha_{0,u}, \ldots, \alpha_{k,u})$ are the linear model coefficients for quantile level u which are estimated by solving

$$\min_{\alpha_{0,u},\ldots,\alpha_{m,u}} \sum_{i,j \le I} \rho_u(\epsilon_{ij}) = \sum_{i,j \le I} \epsilon_{ij}[u - I(\epsilon_{ij} < 0)] \tag{4}$$

and $\epsilon_{ij} = y^*_{ij} - \alpha_{0,u} - \sum\limits_{k=1}^{m} \alpha_{k,u}\, x_{ijk}$. Then the quantile function for the original data is $Q_Y(u|\boldsymbol{x}_{ij}) = \exp(Q_{Y^*}(u|\boldsymbol{x}_{ij}))$. Koenker and Hallock (2001) illustrate the loss function ρ_u for quantile regression as we represent in Figure 3.1.

Koenker and Machado (1999); Yu and Moyeed (2001) show that the solution to minimization of the loss function in equation (4) for estimating the parameter vector $\boldsymbol{\alpha}_u$ is equivalent to maximum likelihood estimation of the parameters of the AL distribution. Hence, the parameter vector $\boldsymbol{\alpha}_u$ can be estimated via an AL distribution with pdf

$$f(y_{ij}^*|\mu_{ij}, \sigma_{ij}^2, p) = \frac{p(1-p)}{\sigma_{ij}} \exp\left(-\frac{(y_{ij}^* - \mu_{ij}^*)}{\sigma_{ij}}[p - I(y_{ij}^* \leq \mu_{ij})]\right) \tag{5}$$

where the skew parameter $0 < p < 1$ gives the quantile level u, $\sigma_{ij} > 0$ is the scale parameter and $-\infty < \mu_{ij}^* < \infty$ is the location parameter. Since the pdf (5) contains the loss function (4), it is clear that parameter estimates which maximize (5) will minimize (4).

In this formulation the AL distribution represents the conditional distribution of the observed dependent variables (responses) given the covariates. More precisely, the location parameter μ_{ij} of the AL distribution links the coefficient vector $\boldsymbol{\alpha}_u$ and associated independent variable covariates in the linear regression model to the location of the AL distribution. It is also worth noting that under this representation it is straightforward to extend the quantile regression model to allow for heteroscedasticity in the response which may vary as a function of the quantile level u under study. To achieve this one can simply add a regression structure linked to the scale parameter σ_{ij} in the same manner as was done for the location parameter.

Equivalently, we assume Y_{ij}^* conditionally follows an AL distribution denoted by $Y_{ij}^* \sim AL(\mu_{ij}^*, \sigma_{ij}^2, u)$. Then

$$Y_{ij}^* = \mu_{ij}^* + \epsilon_{ij}^* \sigma_{ij} \tag{6}$$

where $\epsilon_{ij}^* \sim AL(0, 1, u)$, $\mu_{ij}^* = \alpha_{0,u} + \sum\limits_{k=1}^{m} \alpha_{k,u}\, x_{ijk}$ and $\sigma_{ij}^2 = \exp(\beta_{0,u} + \sum\limits_{k=1}^{\nu} \beta_{k,u}\, s_{ijk})$. Discussion on the choice of link function and structure of regression terms will be undertaken in later sections. In presenting the model in this fashion we already start to move towards the representation of a parametric quantile regression structure.

3.2.2. Parametric Quantile Regression Models. Alternatively, we may adopt a parametric approach to study the quantile regression structure. Two types of distributions, on real support $\Re$ or positive support $\Re^+$ can be considered and we begin with distributions on $\Re$. In this case, we assume that $Y_{ij}^* \sim F(y^*|\boldsymbol{\theta})$ where $F(y^*|\boldsymbol{\theta})$ is the conditional cumulative distribution function (cdf) and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is a vector of model parameters including all unknown coefficient parameters and distributional parameters. The quantile function for the conditional distribution of Y_{ij}^* given $\boldsymbol{x}_{ij}$ at a quantile level $u \in (0, 1)$ is given by:

$$Q_{Y^*}(u|\boldsymbol{x}_{ij}) \equiv \inf\left\{y^* :\; F(y^*|\boldsymbol{\theta}) \geq u\right\}. \tag{7}$$

Under this formulation, the conditional quantile function in (7) can be written as

$$Q_{Y^*}(u|\boldsymbol{x}_{ij}) = \mu_{ij}^* + Q_{\epsilon^*}(u)\sigma_{ij} \tag{8}$$

where $Q_{\epsilon^*}(u) = F_{z^*}^{-1}(u)$ is the inverse cdf for the standardized variable $Z_{ij}^* = \frac{Y_{ij}^* - \mu_{ij}^*}{\sigma_{ij}}$ and again one may incorporate regression structures given as follows for location and scale functions:

$$\textbf{location:} \quad \mu_{ij}^* = \alpha_0 + \sum_{k=1}^{m} \alpha_k x_{ijk}, \tag{9}$$

$$\textbf{scale:} \quad \sigma_{ij}^2 = \exp(\beta_0 + \sum_{k=1}^{\nu} \beta_k s_{ijk}). \tag{10}$$

To transform the quantile function $Q_{Y^*}(u|\boldsymbol{x}_{ij})$ back to the original scale of the data $Y_{ij} = \exp(Y_{ij}^*)$, we suggest $Q_Y(u|\boldsymbol{x}_{ij}) = \exp(Q_{Y^*}(u|\boldsymbol{x}_{ij}))$. We note that there is no unique way to transform the quantile function $Q_{Y^*}(u|\boldsymbol{x}_{ij})$ for Y_{ij}^* back to Y_{ij} and the proposed transformation $Q_Y(u|\boldsymbol{x}_{ij}) = \exp(Q_{Y^*}(u|\boldsymbol{x}_{ij}))$ does not equal in general to the quantile function for the log-AL distribution.

Remark: *We observe that the difference between the non-parametric and the parametric quantile regression models is that in the parametric structure we make explicit the quantile function of the "residual" denoted by $Q_\epsilon(u)$.*

For distributions on $\Re^+$, we assume that $Y_{ij} \sim F(y|\boldsymbol{\theta})$ with mean $\exp(\mu_{ij}^*)$ where μ_{ij}^* is given in (9). Next we make explicit several possible parametric models one may consider in quantile regressions for risk margin. Each model has different associated properties with regard to the relationship of the skewness, kurtosis and heaviness of the tail that it imposes on the quantile function of the response given the covariates.

3.2.2.1. *Asymmetric Laplace Distribution.* As discussed above, the AL distributional family is a useful model structure which naturally fits into a quantile regression framework. As made explicit above, the AL distribution is a three parameter distribution which has been shown to be directly linked to the estimation of quantiles in a quantile regression framework, see further details in Yu and Zhang (2005).

Since this realization, the AL family has been utilized in several financial risk and econometric settings such as Guermat and Harris (2001) who use the symmetric laplace distribution with GARCH volatility to model short-horizon asset returns. Chen et al. (2012) extend this to allow skewness via AL distribution. Yu and Moyeed (2001) apply AL distribution for quantile regression purposes, though as yet, no such developments have been made in the insurance and particularly the risk margin context. Here we propose such a model for risk margin estimation.

If we model the residuals ϵ_{ij} by an AL distribution, the quantile function for observed data Y_{ij}^* is given by (8) where $F_{z^*}^{-1}(u)$ is the inverse cdf (quantile function)

$$F_{AL}^{-1}(u|\mu, \sigma^2, p) = \begin{cases} \mu + \frac{\sigma}{1-p} \log(\frac{u}{p}), & \text{if } 0 \leq u \leq p, \\ \mu - \frac{\sigma}{p} \log(\frac{1-u}{1-p}), & \text{if } p < u \leq 1. \end{cases} \tag{11}$$

To understand how the three location, shape and scale parameters of the AL distribution affect the shape and tails of the distribution it is also useful to note the following relationship between the parameters and the mean, variance, skewness S and kurtosis K of AL distribution:

$$E(Y) = \mu + \frac{\sigma(1-2p)}{p(1-p)}, \qquad Var(Y) = \frac{\sigma^2(1-2p+2p^2)}{(1-p)^2p^2}, \tag{12}$$

$$S(Y) = \frac{2[(1-p)^3 - p^3]}{((1-p)^2+p^2)^{3/2}}, \qquad K(Y) = \frac{9p^4 + 6p^2(1-p)^2 + 9(1-p)^4}{(1-2p+2p^2)^2}. \tag{13}$$

Note that the shape parameter p of the AL distribution gives the magnitude and direction of skewness. AL distribution is skewed to left when $p > 0.5$ and skewed to right when $p < 0.5$ and hence it can model the left skewness of most log transformed loss data directly through this shape parameter p. Moreover as the risk margin adopted in insurance industry is mostly greater than 50 percent, AL distribution allows the calculation of quantiles rather than mean estimates fairly easily. Figures 3.2(a) and 3.2(b) show a variety of pdf for AL distribution and its skewness and kurtosis respectively.

FIGURE 3.2. (a) The pdf of asymmetric Laplace distribution

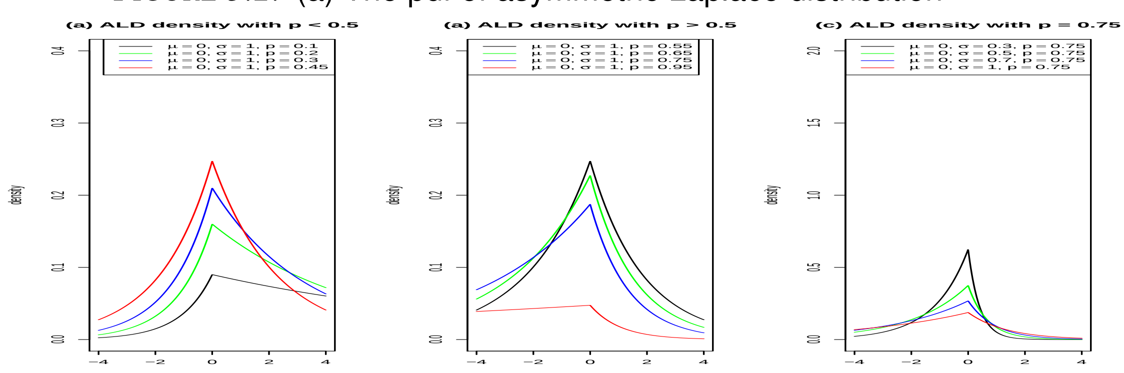
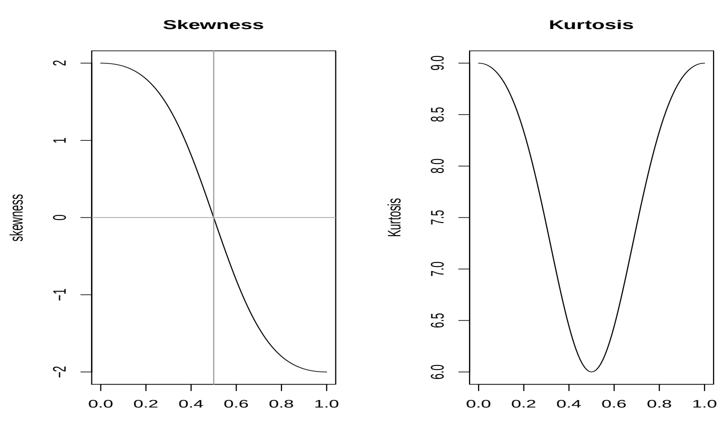


Figure 2: (b) The skewness and kurtosis of asymmetric Laplace distribution



3.2.2.2. *Power Pareto Model.* As the second choice of parametric quantile regression model we consider the framework of **?**. In this approach a polynomial power-Pareto (PP) quantile function model is developed. This model combines a power distribution with a Pareto distribution, which enables us to model both the main body and the tails of a distribution. In considering the PP model the conditional quantile function of the response (reserve in each cell) are comprised of two components:

- component 1: a power distribution $F_1(y) = y^{\frac{1}{\gamma_1}}$ where $y \in [0,1]$ and $\gamma_1 > 0$ with a corresponding quantile function then given by $Q_1(u;\gamma_1) = u^{\gamma_1}$ for $u \in [0,1]$; and
- component 2: a Pareto distribution function $F_2(y) = 1 - y^{-\frac{1}{\gamma_2}}$ where $y \geq 1$ and $\gamma_2 > 0$ with a corresponding quantile function then given by $Q_2(u;\gamma_2) = (1-u)^{-\gamma_2}$.

One may use the fact that the product of the two quantile functions will remain a strictly valid quantile function producing the new quantile function family known as the Polynomial-Power Pareto model. The resulting structural form given by the inverse cdf of the Pareto distribution with an additional polynomial power term:

$$F^{-1}_{PP}(u|\gamma_1,\gamma_2) = u^{\gamma_1}(1-u)^{-\gamma_2}. \tag{14}$$

Hence the quantile function is again given by (8) where $Q_{\epsilon^*}(u) = F^{-1}_{PP}(u)$ and $Q_Y(u) = \exp(Q_{Y^*}(u))$.

From the specification of this quantile function, one may then derive the resulting pdf of the PP model for $Y^*_{ij} = \ln Y_{ij}$ which is given by

$$f_{PP}(y^*_{ij}|\gamma_1,\gamma_2) = \frac{u_{ij}^{1-\gamma_1}(1-u_{ij})^{\gamma_2+1}}{\sigma_{ij}[\gamma_2 u_{ij} + \gamma_1(1-u_{ij})]}$$

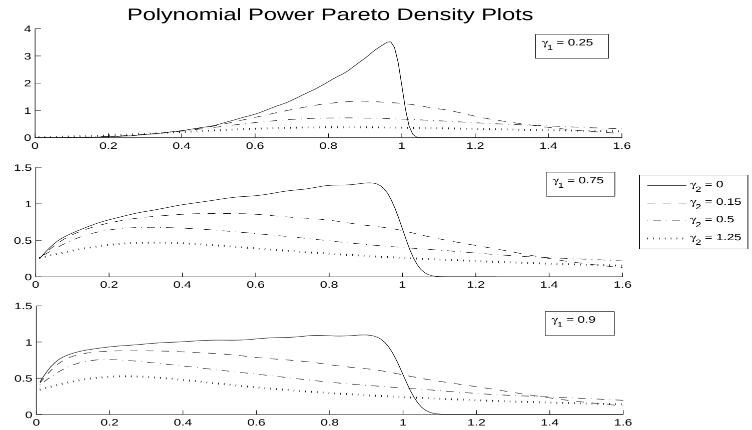
where u_{ij} is given by solving the system of equations defined for each observation by

$$y^*_{ij} = \mu^*_{ij} + u_{ij}^{\gamma_1}\left(1-u_{ij}\right)^{-\gamma_2}\sigma_{ij}. \tag{15}$$

where again we treat the location $\mu^*_{ij} = \mu^*_{ij}(\boldsymbol{\alpha})$ in (9) and scale $\sigma_{ij} = \sigma_{ij}(\boldsymbol{\beta})$ in (10) as functions of the regression coefficients and associated covariates. We note that in this case the u_{ij} is really an implicit function of the regression structure as each u_{ij} is found as the solution to the system of equations in (15).

To complete the specification of the polynomial power Pareto model we plot the shape of the density that can be obtained for a range of different power parameters for the power and pareto components with a unit scale factor $\sigma = 1$. These plots in Figure 3.3 demonstrate the flexible skewness, kurtosis and tail features that can be obtained from such a model by varying the parameters γ_1 and γ_2.

FIGURE 3.3. The pdf of Power Pareto distribution



3.2.2.3. *Generalised Beta Distribution of the Second Type Family.* We note that the AL and PP families of quantile regression models require a log transformation of the data before the modelling to ensure the data has real support $\Re$ that these distributions are defined upon. In performing this transformation, one must analyze carefully the effect of the transformation on the ability to fit such models and the resulting model interpretability must be interpreted with regard to the transformation. This is particularly the case if zero counts are present in the data for some accident and development years. Moreover, in the context of claims reserving, loss data often exhibits heavy-tailed behavior, particularly for long tail business classes. To account for such features and to remove the need to consider pre-transformation of the data one may consider the family of generalized beta (GB2) distributions of the second kind.

The type two generalized beta distribution (GB2) has attractive features for modelling loss reserve data, as it has a positive support $\Re^+$ and nests a number of important distributions as its special cases. The GB2 distribution has four parameters, which allows it to be expressed in various flexible densities. See Dong and Chan (2013) for a more detailed description of GB2 distribution including its pdf and distribution family.

If $Y_{ij} \in \Re^+$ conditionally follows a GB2 distribution, then it can be characterized by the density given by

$$f_{GB2}(y_{ij}|a, b_{ij}, p, q) = \frac{\frac{a}{b_{ij}}\left(\frac{y_{ij}}{b_{ij}}\right)^{ap-1}}{B(p,q)[1 + (\frac{y_{ij}}{b_{ij}})^a]^{p+q}}, \quad \text{for } y_{ij} \geq 0 \tag{16}$$

where a, p and q are shape parameters and b_{ij} is the scale parameter.

In particular, b_{ij} can be linked to the mean μ_{ij} of the distribution as follows:

$$b_{ij} = \frac{\mu_{ij} B(p,q)}{B(p + 1/a, q - 1/a)} \tag{17}$$

where μ_{ij} is log-linked to a linear function of covariates μ^*_{ij} in (9) according to the relationship:

$$E(Y_{ij}) = \mu_{ij} = \exp\left(\alpha_0 + \sum_{k=1}^{m} \alpha_k\, x_{ijk}\right). \tag{18}$$

Then the variance is given by:

$$Var(Y_{ij}) = \mu_{ij}^2 \left\{ \frac{B(p,q)B(p+2/a, q-2/a)}{[B(p+1/a, q-1/a)]^2} - 1 \right\}. \tag{19}$$

The GB2 distribution is a generalization from the beta distribution with pdf:

$$f_B(z_{ij}|p,q) = \frac{1}{B(p,q)} z_{ij}^{p-1}(1-z_{ij})^{p+q} \tag{20}$$

via the transformation $z_{ij} = \dfrac{(\frac{y_{ij}}{b_{ij}})^a}{1+(\frac{y_{ij}}{b_{ij}})^a}$. Hence the cdf of GB2 distribution is given by:

$$F_{GB2}(y_{ij}|a, b_{ij}, p, q) = \int_0^{z_{ij}} \frac{t^{p-1}(1-t)^{(q-1)}}{B(p,q)} dt = \frac{B(z_{ij}|p,q)}{B(p,q)} = F_B(z_{ij}|p,q) \tag{21}$$

where $B(z_{ij}|p,q)$ is the incomplete beta function.

The GB2 is directly relevant for quantile regression models since one may also find its quantile function in closed form according to the following expression:

$$Q_Y(u) = \frac{\exp\left(\alpha_0 + \sum\limits_{k=1}^{m} \alpha_k x_{ijk}\right) B(p,q)}{B(p+1/a, q-1/a)} \left(\frac{F_B^{-1}(u|p,q)}{1-F_B^{-1}(u|p,q)}\right)^{\frac{1}{a}}. \tag{22}$$

There are many widely known and utilized sub-families of the GB2 family, we present two examples of relevance to the context of risk margin estimation that we will explore, corresponding to the generalized gamma and the gamma distribution sub-families.

3.2.2.4. *Two Special Cases of GB2.* To understand the flexibility of the GB2 family of models, we consider the case when $q = \infty$, then the resulting GB2 distribution sub-family becomes the generalized gamma (GG) distribution, see discussion in McDonald (1984). The GG family of models was independently introduced by Stacy (1962), as a three parameter distribution with pdf given by:

$$f_{GG}(y_{ij}|a, b_{ij}, p) = \lim_{q\to\infty} \frac{\frac{a}{b_{ij}}(\frac{y_{ij}}{b_{ij}})^{ap-1}}{B(p,q)[1+(\frac{y_{ij}}{b_{ij}})^a]^{p+q}} = \frac{a(\frac{y_{ij}}{b_{ij}})^{ap}\exp[-(\frac{y_{ij}}{b_{ij}})^a]}{y_{ij}\Gamma(p)}, \quad \text{for } y_{ij} > 0 \tag{23}$$

where a and p are shape parameters and b_{ij} is scale parameter linked to the mean of the distribution as:

$$b_{ij} = \frac{\mu_{ij}\Gamma(p)}{\Gamma(p+1/a)} \tag{24}$$

and the mean is again log-linked to a linear function of covariates in (18). The cdf is

$$F_{GG}(y_{ij}|a, b_{ij}, p) = \int_0^{z_{ij}} \frac{t^{p-1}e^{-t}}{\Gamma(p)} dt = \frac{\gamma_1(z_{ij}|p)}{\Gamma(p)} = F_G(z_{ij}|1,p)$$

where $\gamma_1(z_{ij}|p)$ is the lower incomplete gamma function and $z_{ij} = (\frac{y_{ij}}{b_{ij}})^a$. Hence, the quantile function is given by:

$$Q_Y(u) = \frac{\exp\left(\alpha_0 + \sum_{k=1}^{m} \alpha_k x_{ijk}\right)\Gamma(p)}{\Gamma(p+1/a)} \left(F_G^{-1}(u|1,p)\right)^{\frac{1}{a}} \tag{25}$$

The second case is nested within the GG family and corresponds to the two parameter Gamma distribution which is obtained by further restricting $a = 1$. Its pdf and quantile function are well-known and can be expressed using equations (23) and (25) by replacing a with 1.

Having defined clearly the three different quantile regression distributional families that will be considered in the parametric quantile regression framework, we now introduce the different regression structures we consider in the quantile regression under each distributional assumption.

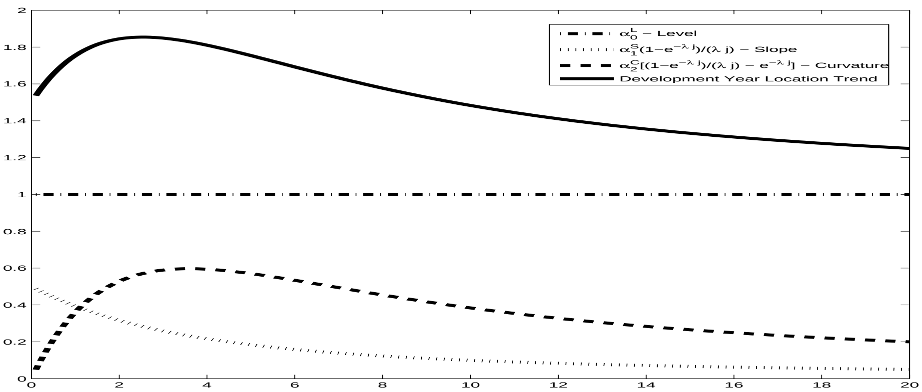
3.2.3. Structural Components of the Quantile Regression Framework. In the model structures we will adopt, as is standard practice in regression modelling, once we believe we have suitable explanatory variables for the dependent variable quantity of interest, in this case the conditional quantile function, we will assume the observations are independent.

In the following subsections we explain how under each different distributional assumption for the conditional quantile regression structure, one may introduce a link function to relate regression models using independent covariates to the response quantiles in order to model trend behaviors in the location and scale of the quantile function. To simplify all the possible different model considerations we consider only log link functions in all regressions.

The possible regression structures we consider will be classified as: location based explanatory factors i.e. trends in accident and development years; and scale (heteroskedascity / variance) based explanatory factors for accident and development years. We note that when it comes to different distributional choices since we may transform the observations, we are actually considering both additive and multiplicative (mixed interaction) terms in our regressions and as such we explore aspects of ANOVA as well as ANCOVA regression structures in the quantile regression setting. A summary of the model structures we consider for the location and scale components of each model is provided in Table 3.7 in Appendix 1. We note that in general one may consider that a version of the ANCOVA model was applied to the PP and AL models and a version of the ANOVA model was effectively applied to the AL and GB2 families. In addition we will allow the influence of covariates to affect different quantile levels to different extents, making for an interesting analysis on the effect of model structure on quantile level.

We note that since the focus of this manuscript differs to that undertaken in the Poisson-Tweedie regression context of Peters et al. (2009), in that the focus of the regression model comparison

FIGURE 3.4. Basis function regression structure for development years in location parameter in the AL model ($M_{1\cdot}$)



will be primarily concerned with the model choice for the distributional form of the conditional quantile function, not so much on the model structure uncertainty related to all possible covariate model sub-space structures and nested models, therefore we limit the analysis to the ANOVA and ANCOVA structures given below. If one is interested in specialized techniques to explore and compare all possible models sub-spaces within each distributional model, we suggest the approach adopted in Peters et al. (2009), or recently in Verrall and Wuthrich (2014).

3.2.3.1. *Location: Development and Accident Year Trend Model Structures.* The primary sets of covariates we consider correspond to the accident year and the development year in the claims reserving structure, as well as transformations of these through basis functions. From Table 3.7 one may observe that we label models using two subscripts according to their mean and variance functions respectively. Models 0• (denoted by $M_{0\cdot}$) and 1• (denoted by $M_{1\cdot}$) are parsimonious location structure specifications for the general model in (9) with $m = 2$, that is, the additive structure is given by:

$$\text{Model } 0\bullet: \quad \mu_{ij}^* = \alpha_0 + \alpha_1 \times i + \alpha_2 \times j, \tag{26}$$

$$\text{Model } 1\bullet: \quad \mu_{ij}^* = \alpha_0 + \alpha_1^S F_1(j) + \alpha_2^C F_2(j). \tag{27}$$

Under $M_{0\cdot}$ one assumes a linear trend across accident and development years. If a non-linear trend across development years is considered with an assumption of common behavior down the accident years, on may consider $M_{1\cdot}$ which is a basis regression model popular in term structure models and known as the Nelson-Seigel model (Nelson and Siegel, 1987). Examples of typical basis functions we considered under this choice for the location are given in Figure 3.4 below, where we show the 'level', 'slope' and 'curvature' structure of the location trend from such a model. In the graph, we show the decomposition of the role the level, slope and curvature basis functions which play in the regression with example coefficients: $\alpha_0 = 1$, $\alpha_1^S = 0.5$, $\alpha_2^C = 2$ and $\lambda = 0.5$ with $j \in \{1, 2 \ldots, J\}$ in years.

In the context of an ANOVA model specification for the location one can assume a form given by:

$$\text{Model 2}\bullet\text{:} \quad \mu^*_{ij} \;=\; \alpha_0 + \alpha_{1i} + \alpha_{2j}. \tag{28}$$

This location (trend) function corresponds to the general model in (9) with $m = 2$,

$$\alpha_1 x_{ij1} = \alpha_{1i} \quad \text{and} \quad \alpha_2 x_{ij2} = \alpha_{2j}.$$

The parameters α_{1i} and α_{2j} denote the accident year and development year effects respectively and they satisfy the following constraints:

$$\alpha_{11} = \alpha_{21} = 0. \tag{29}$$

This parametrization is set up in the context of loss reserving so that all parameters are relative to the first accident year which has the most information. These location functions (26) to (28) apply to both AL and PP distributions in general. For Gamma, GG and GB2 distributions with positive support $\Re^+$, a log link function is considered and the location functions become $\mu_{ij} = \exp(\mu^*_{ij})$. When the AL distribution, with the shape parameter $p = u$ is applied, Model 3• ($M_{3\cdot}$) corresponds a nonparametric quantile function

$$\text{Model 3}\bullet\text{:} \quad \mu^*_{ij,u} \;=\; \alpha_{0,u} + \alpha_{1i,u} + \alpha_{2j,u} \tag{30}$$

where $\alpha_{\bullet,u}$ are parameters at quantile level u.

3.2.3.2. *Scale: Development and Accident Year Variance Model Structures.*

There are different choices for the structure of the variance function for the AL and PP distributions but Gamma, GG and GB2 distributions do not have a component to model σ^2 directly. Model •0 ($M_{\cdot 0}$) assumes homoscedastic variance $\sigma^2_{ij} = \sigma^2$. Models •0 ($M_{\cdot 0}$) to •3 ($M_{\cdot 3}$) are specified below:

$$\text{Model } \bullet 0: \quad \sigma^2_{ij} \;=\; \sigma^2, \tag{31}$$

$$\text{Model } \bullet 1: \quad \sigma^2_{ij} \;=\; \exp(\beta_0 + \beta_{1i}), \tag{32}$$

$$\text{Model } \bullet 2: \quad \sigma^2_{ij} \;=\; \exp(\beta_0 + \beta_{2j}), \tag{33}$$

$$\text{Model } \bullet 3: \quad \sigma^2_{ij} \;=\; \exp(\beta_0 + \beta_{1i} + \beta_{2j}), \tag{34}$$

where the parameters β_{1i} and β_{2j} which denote the accident year and development year effects respectively satisfy the following constraints:

$$\beta_{11} = \beta_{21} = 0. \tag{35}$$

Again Models •1 to •2 corresponds to (10) with $\beta_1 s_{ij1} = \beta_{1i}$ and $\beta_2 s_{ij2} = \beta_{2j}$. Furthermore, for Model 23', the shape parameter in the AL distribution is further modelled by the accident year effect, which is specified as follows:

$$\text{Model 23'}: \quad p_i \;=\; \phi_0 + \phi_{1i}. \tag{36}$$

where the parameters ϕ_{1i} denote the accident year effect and satisfy the following constraints:

$$\phi_{11} = 0. \tag{37}$$

3.3. Bayesian Framework: Posterior Quantile Regression

The estimation of quantile regression models is straightforward to adopt under a Bayesian formulation. One of the key advantage of using Bayesian procedures for practical models such as those we develop above lies in the adoption of available prior information and the provision of a complete predictive distribution for the required reserves (De Alba, 2002).

To complete the posterior distribution specification in each model it suffices to consider the representation of two components: the likelihood of the data for the regression structure (that is, the density not the quantile function); and the prior specifications for the model parameters. In the above sections, the quantile function of the likelihood is presented, along with the associated density for the observations conditional upon the parameters and covariates, that is, the likelihood for each model. Therefore, to formulate the Bayesian structure we simply need to present the prior structures we consider for the parameters in each model. This will be relatively straightforward for models formed from the AL distribution structure and the GB2 structures, but not so trivial for the case of the PP model.

In the real data examples we consider below, we adopt an objective Bayesian perspective in which we consider relatively uninformative priors. This reflects our lack of prior knowledge for the model parameters likely ranges or magnitudes. For instance, the priors for parameters (coefficients) in mean, variance and skewness quantile regression functions are all selected as Gaussian:

$$\alpha_0,\ \alpha_1,\ \alpha_1^S,\ \alpha_{1i},\ \alpha_2,\ \alpha_2^C,\ \alpha_{2j},\ \beta_{1i},\ \beta_{2j},\ \phi_0,\ \phi_{1i} \sim N(0,100) \tag{38}$$

and for the shape parameters of the GB2 distribution are:

$$a \sim N(0,100), \qquad p \sim Ga(0.001, 0.001), \qquad q \sim Ga(0.001, 0.001). \tag{39}$$

Normal and gamma distributions are standard choices of priors for parameters with a real and positive support respectively, see discussions on possible choices in Denison et al. (2002). In the case of the AL and GB2 models, these priors combined with the resulting likelihoods produce in each case standard and well defined posterior distributions.

In the case of the PP model one has to be careful to define the posterior support to ensure the resulting distribution is normalized and therefore a proper posterior density. To ensure this is the case one must impose constraints on the posterior support which can be uniquely characterized by the three sets of parameter space constraints Ω_1, Ω_2 and Ω_3, for coefficient vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and

(γ_1, γ_2) respectively, given by:

$$\begin{aligned}
\Omega_1 &= \left\{ (\alpha_{0,u}, \dots, \alpha_{m,u}) \ : \ \alpha_{0,u} + \sum_{k=1}^{m} \alpha_{k,u} x_{ijk} < y_{ij}, \ \ \forall i,j \in \{1,2,\dots,I\} \right\}, \\
\Omega_2 &= \left\{ (\beta_{0,u}, \dots, \beta_{\nu,u}) \ : \ \beta_{0,u} + \sum_{k=1}^{\nu} \beta_{k,u} s_{ijk} > \epsilon > 0, \ \ \forall i,j \in \{1,2,\dots,I\} \right\}, \\
\Omega_3 &= (0, M] \times (0, \infty), \ \ M \in \Re^{+}.
\end{aligned} \tag{40}$$

Under these parameter space restrictions the resulting posterior for the PP model can be shown to be well defined as a proper density, see a derivation and proof in Theorem 1 of **?**.

In **?** they consider an MCMC scheme for the resulting posteriors based on standard Metropolis-Hastings steps with rejection when the proposed parameter values fail to satisfy the posterior support constraints. In general this results in a very slowly mixing MCMC chain which will have very poor properties. We replace this idea with simple block Metropolis within Gibbs updates which allow for smaller moves in each component of the constrained posterior support making it more likely to satisfy the constraints and also simpler to design and tune the proposal for the MCMC scheme. This was a significant improvement compared to the approach proposed in **?**. We implement this sampler in `R`. For the other Bayesian models from the AL and GB2 models, sampling from the intractable posterior distributions involved the Gibbs sampling algorithm (Smith and Roberts, 1993; Gilks et al., 1996) and Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953) are the most popular MCMC techniques. For readers who are less familiar with Bayesian computation techniques, we suggest the `WinBUGS` (Bayesian analysis Using Gibbs Sampling) package, see Spiegelhalter et al. (2002). The MCMC algorithms that are implemented for each model in `WinBugs` and `R` are available upon request.

In the Gibbs sampling scheme, a single Markov chain is run for 60,000 to 110,000 iterations, discarding the initial 10,000 iterations as the burn-in period to ensure convergence of parameter estimates. Convergence is also carefully checked by the history and autocorrelation function (ACF) plots. The every 10-th simulated values from the Gibbs sampler after the burn-in period are sampled to mimic a random sample of size 5000 to 10,000 from the joint posterior distribution for posterior inference. Parameter estimates are given by the posterior means.

3.4. Quantile Prediction for Risk Measures, Risk Margin

As discussed in the introduction, the predicted reserves are typically performed in a claims reserving setting by predicting the mean reserve in each cell in the lower triangle $\mathcal{D}_l$. Other methods for reserving may involve the quantification of a risk measure based on the distribution of the predicted reserves, in place of the mean predicted reserve, such as VaR, Expected Shortfall (ES) or Spectral Risk Measures (SRM), see discussions in the tutorial review of Peters et al. (2013). In addition, in order to quantify the uncertainty in a central measure for the predicted reserve, one may alternatively take the central measure of reserve and make a risk margin adjustment based on the distribution of the predicted reserves in the form of a quantile function.

When calculating any of these required measures for the resulting total outstanding reserves one requires to first obtain the predictive density, which under the Bayesian setting can be obtained for instance in one of the following two ways for each $Y_{ij} \in \mathcal{D}_l$:

- **Full Predictive Posterior Distribution:**
$$F_{Y_{ij}}\left(y_{ij}|\mathcal{D}_0\right) = \int_0^{y_{ij}} f_{Y_{ij}}\left(y|\mathcal{D}_0\right)\, dy = \int_0^{y_{ij}} \int f_{Y_{ij}}\left(y|\boldsymbol{\theta}\right) \pi\left(\boldsymbol{\theta}|\mathcal{D}_0\right)\, d\boldsymbol{\theta}\, dy.$$
Here, all posterior parameter uncertainty is integrated out of the predictive distribution.
- **Conditional Predictive Posterior Distribution:**
$$F_{Y_{ij}}\left(y_{ij}|\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)\right) = \int_0^{y_{ij}} f_{Y_{ij}}\left(y|\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)\right)\, dy$$
where the point estimator $\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)$ contains the information from the upper triangle. Examples of common estimators include the posterior mean $\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right) = \widehat{\boldsymbol{\theta}}^{(MMSE)}$ or mode $\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right) = \widehat{\boldsymbol{\theta}}^{(MAP)}$.

Using these predictive distributions, one may also be interested in quantities such as the distribution of the total outstanding claim given by the sum of the losses in the lower triangle according to the random variable $Y_T := \sum_{(i,j)\in\mathcal{D}_l} Y_{ij}$ which has distribution given under the full predictive posterior distribution according to convolution given by

$$\begin{aligned} F_{Y_T}\left(y_t|\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)\right) &:= *_{(i,j)\in\mathcal{D}_l} F_{Y_{ij}}\left(y|\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)\right) \\ &= \left(F_{Y_{I,2}} * F_{Y_{I-1,3}} * F_{Y_{I-2,4}} * \cdots * F_{Y_{I,I}}\right)\left(y|\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)\right). \end{aligned} \tag{41}$$

where, one convolves the distributions for the loss elements in the lower triangle with $*$ the convolution operator. One can then state several features about the tail behavior of the total loss distribution and also therefore of the high quantiles as $y \to \infty$, depending on the properties of the individual loss random variables in the sum. For instance, if one has loss distributions on $\Re^+$ then one can obtain the lower bound given by

$$\begin{aligned} \overline{F_{Y_T}}\left(y_t|\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)\right) &:= \overline{\left(F_{Y_{I,2}} * F_{Y_{I-1,3}} * F_{Y_{I-2,4}} * \cdots * F_{Y_{I,I}}\right)}\left(y|\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)\right) \\ &\sim c \sum_{(i,j)\in\mathcal{D}_l} \overline{F_{Y_{ij}}}\left(y|\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)\right), \quad \text{as } y \to \infty, \end{aligned} \tag{42}$$

for some $c \geq 1$. Note, if at least one of the lower triangle losses Y_{ij} is distributed according to a heavy tailed loss distribution, such as sub-exponential, regularly varying or long tailed loss distributions then one can find the precise value for c. For instance if the total loss is max-sum equivalent, then $c = 1$, see definitions for regular variation, sub-exponential, long tailed and max-sum equivalence in Bingham et al. (1989) and in the context of insurance and quantile function approximations as discussed here, see the recent tutorial and references therein from Peters et al. (2013).

These conditional predictive distributions can be obtained for any model approximately by solving the integrals using the Markov chain Monte Carlo samples obtained from the posterior $\pi\left(\boldsymbol{\theta}|\mathcal{D}_0\right)$. Then, given a predictive distribution, one can then find quantile functions according to the following approaches:

- **Full Predictive Posterior Quantile Function:** is given by $Q_{Y_{ij}|\mathcal{D}_0}(u) := F^{-1}_{Y_{ij}}\left(y_{ij}|\mathcal{D}_0\right)$ which is the solution to the second order ordinary differential equation:
$$\frac{d}{dQ_{Y_{ij}|\mathcal{D}_0}} f_{Y_{ij}}\left(Q_{Y_{ij}|\mathcal{D}_0}(u)\,|\mathcal{D}_0\right)\left(\frac{dQ_{Y_{ij}|\mathcal{D}_0}}{du}\right)^2 + f_{Y_{ij}}\left(Q_{Y_{ij}|\mathcal{D}_0}(u)\,|\mathcal{D}_0\right)\frac{d^2 Q_{Y_{ij}|\mathcal{D}_0}}{du^2} = 0,$$
which is obtained by twice differentiating the following identity:
$$F_{Y_{ij}}\left(Q_{Y_{ij}|\mathcal{D}_0}(u)\,|\mathcal{D}_0\right) = \int_0^{Q_{Y_{ij}|\mathcal{D}_0}(u)} f_{Y_{ij}}\left(y|\mathcal{D}_0\right) dy = u. \tag{43}$$
The solution to this second order ordinary differential equation can often be found in the form of a power series, see discussions in Gyorgy and Shaw (2008).
- **Conditional Predictive Posterior Quantile Function:**
$$Q_{Y_{ij}|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)}(u) := F^{-1}_{Y_{ij}}\left(u|\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)\right) \tag{44}$$
which is the most convenient choice that we recommend since the inverse of the predictive distribution in this case takes the closed form expressions for the particular model considered as detailed in Section 2.2.
- **Conditional Total Reserve Posterior Quantile Function:** In many cases one is also interested in finding the quantile function of the distribution corresponding to the total reserve, which under conditional independence is given by $F^{-1}_{Y_T}\left(y_t|\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)\right)$ where this is given by the quantile function of the distribution in Equation 41. In general finding the convolution and inverse of this convolved distribution must be done numerically. There are many basic results known about these quantities such as asymptotic results and bounds for different properties of light and heavy tailed random variables, independent or dependent, see a discussion in Kaas et al. (2000).

  Light Tailed Run-off for Claims Process: In the case in which no loss cells in the claims triangle are heavy tailed, then in general one would need to approximate the tail quantile for the partial sum of all losses. In Kaas et al. (2000) they study partial sums of random variables with no assumption of independence or of identical marginal distributions. The only assumption is that the tails are not so heavy for each marginal, such that each marginal has finite mean. It will be useful to recall that for two random variables X and Y, X proceeds Y under convex ordering $X \leq_{CX} Y$ iff for all convex real functions $g(\cdot)$ with finite expectations one has

$$\mathbb{E}\left[g(X)\right] \leq \mathbb{E}\left[g(Y)\right]. \tag{45}$$

Thus, two random variables X and Y with equal mean are convex ordered if their cdfs cross once.

Then one can show that in such cases for any sequence of loss distributions $\left\{F_{Y_{ij}}\right\}_{(i,j)\in\mathcal{D}_l}$ the following convex order relationship holds

$$\sum_{(i,j)\in\mathcal{D}_l} Y_{ij} \leq_{CX} \sum_{(i,j)\in\mathcal{D}_l} F^{-1}_{Y_{ij}}(U) \tag{46}$$

for $U \sim U[0,1]$, see derivations in Goovaerts et al. (2000). This result means that the total loss Y_T in the convex order sense, comprised of the most risky joint vector of losses with given marginals, has the comonotonous joint distribution. The components of which are maximally dependent since all components are non-decreasing functions of a common random variable U.

Hence, we consider the following quantile function approximation for the total loss based on the most conservative estimate using the above bound, given by

$$F^{-1}_{Y_T}(u) = \sum_{(i,j)\in\mathcal{D}_l} F^{-1}_{Y_{ij}}(u). \tag{47}$$

Note, in the case of heavy tailed losses this can be refined for large quantiles as follows.

Heavy Tailed Run-off For Claims Process: Alternatively, if additional features of the loss distributions in the lower triangle are known, such as these loss models contain at least one heavy tailed loss distribution, then one can bound the total quantile function result. This can be done conservatively by instead considering the $\mathcal{T}$-fold convolution of the distribution, say $F^{(*\mathcal{T})}_{Y_{i*j*}}$ which correspond to the loss distribution amongst all the lower trianglular loss elements with the dominant index of regular variation (that is, with the heaviest tails). In such cases it would be popular to utilize an asymptotic result for the quantile function of the sum, as the quantile level becomes large $u \to 1$. For instance, one could use the first order or second order asymptotic results, see discussions in Peters et al. (2013); Cruz et al. (2014). As an example, if the quantile regression was structured such that the distribution of the partial sum $Y_T = \sum_{(i,j)\in\mathcal{D}_l} Y_{ij} \sim F_{Y_T}$ is regularly varying with index $\rho \geq 0$ with conditionally i.i.d. Y_{ij} with each Y_{ij} taking positive support, then one can write the first order tail approximation which is asymptotically equivalent to the following

$$\overline{F}_{Y_T}(y) \sim \mathcal{T}\overline{F}_{Y_{i*j*}}(y), \quad y \to \infty, \tag{48}$$

see detailed tutorial in Peters et al. (2013). This would lead to the approximation of the required quantile asymptotically by the expression

$$
\begin{aligned}
Q_{Y_T|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)}(u) &:= \inf\left\{y \in \mathbb{R}^+ :\ F_{Y_T}(y) > u\right\} \\
&\approx \inf\left\{y \in \mathbb{R}^+ :\ \mathcal{T}\overline{F}_{Y_{i*j*}}(y) < 1-u\right\} \\
&\approx Q_{Y_{i*j*}|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)}\left(1-\frac{1-u}{\mathcal{T}}\right) := F^{-1}_{Y_{i*j*}}\left(1-\frac{1-u}{\mathcal{T}}|\widehat{\boldsymbol{\theta}}\left(\mathcal{D}_0\right)\right)
\end{aligned}
\tag{49}
$$

3.5. Model Structure Analysis for Israel data

In this section we perform two core studies: The first involves isolating the structural components for the quantile regressions, in order to perform a study on the mean function and variance functions that are most suitable for an example of a representative claims reserving data set. This is therefore performed using the non-parametric and Bayesian formulations of the AL model with different assumptions on the mean and variance functions. The second involves isolating the distributional choices of the quantile regression, where we take the best fitting parametric model mean and variance function structures and use these to study distributional properties under the different quantile function choices.

The data set used throughout this section is interesting for such a benchmark exercise as it has been previously studied and its features are reasonably well known, see Chan et al. (2008) for more details on the Israel Data set. The data is available in Figure 3.18 in Appendix I and represents the paid out claim amounts y_{ij} for an Israel insurance company, covering periods from 1978 to 1995, containing 171 observations. For mathematical convenience, two zero claim amounts have been replaced with 0.01. Some general trends are observed in this data. Given an accident year, the claim development amounts generally increase between the first 4 to 6 development years then this increase is followed by a generally decreasing trend thereafter. The mean, median, variance and kurtosis of this data are 4459.7, 3,871, 12,059,232.6 and -0.4 respectively. The overall skewness is 0.58 and on a log scale is -2.67.

This data has been studied in Chan et al. (2008) using the generalized-t (GT) distribution expressed as scale mixtures of uniform which facilitates the Bayesian implementation. They adopt the ANOVA and ANCOVA mean structures to study the accident year and development year effects on the conditional mean functions but not on any quantile level. Moreover they also remark that the log transformed data become negatively skewed which the symmetric GT distribution fails to accommodate. Hence, they suggest to adopt some skewed error distributions to improve the model performance.

Our primary point of departure for these previous studies on this data is the conjecture that using a measure of average effects may not be appropriate for understanding loss reserves at higher quantiles. Higher quantile projection is critical in loss reserving, for reinsurance premium calculations and also in deriving the risk margin. In this section, we use all the models in Section 2 for quantile projection with an aim to provide a more comprehensive study on model performance with a wide range of distributions having different tails behavior and model structures for the quantile trends and heteroscedasticity in the accident and development years.

3.5.1. Analysis of Quantile Regression Models: Location and Scale. To investigate the model structures for location (mean) and scale (variance) functions, we consider two settings: the first class of models involves the parametric models using the AL distribution with p either fixed (denoted by fix) or left to be estimated (denoted by est), the mean functions given by (26) to (28) and variance being constant (Models 00-20) or given by (34) (Models 03-23); the second class of models involves a set of nonparametric models which are also studied with mean function (28) and variance being constant or given by (34) (Models 30 and 33) using AL as a proxy distribution with p fixed at different quantile levels.

For model comparison, deviance information criterion (DIC) is adopted, see Appendix III for details. Since, models with smaller DIC are preferred to those with larger DIC, then the results of the model comparisons provided in Table 3.1 show that among the parametric models, M_{23} which incorporates an ANOVA model for both accident and development years in modelling both the mean and variance functions is the best fitting model according to DIC. This show that the accident year and development year effects are both important in describing the dynamics of the mean and variance. Hence, these ANOVA-type mean and variance functions are applied to most of the subsequent analyses whenever possible. For the nonparametric models, M_{33} with ANOVA variance provide better fit than M_{30} with constant variance.

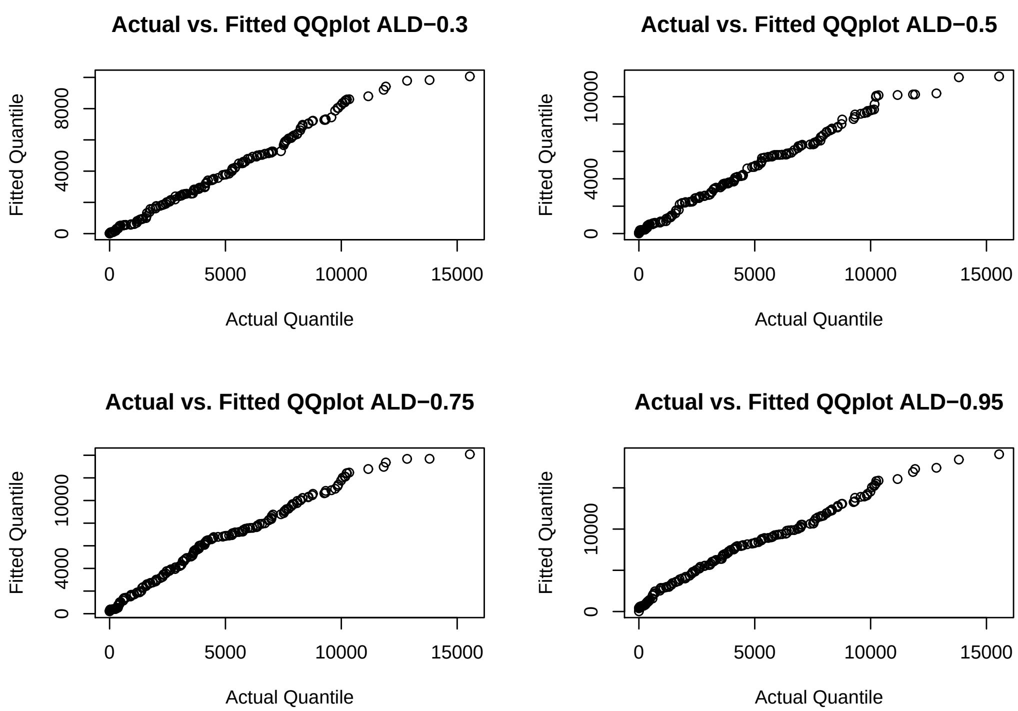
TABLE 3.1. Estimates of p and model fit measures for AL parametric and non-parametric models

Models	DIC	$\bar{D}^\dagger$	$\hat{D}^\ddagger$	p	Models	DIC	$\bar{D}^\dagger$	$\hat{D}^\ddagger$	p
	Variance Constant					Variance Function			
M_{00}	195.41	255.21	315.02	0.85 (est)	M_{03}	272.82	334.74	396.66	0.93 (est)
M_{10}	223.30	284.10	344.91	0.88 (est)	M_{13}	199.14	247.49	295.85	0.95 (est)
M_{20}	50.94	120.17	189.40	0.81 (est)	M_{23}	-20.81	24.91	70.63	0.75 (est)
M_{30}	55.94	125.61	195.28	0.30 (fix)	M_{33}	-37.06	38.34	113.74	0.30 (fix)
M_{30}	73.10	152.26	231.43	0.50 (fix)	M_{33}	-38.80	35.51	109.82	0.50 (fix)
M_{30}	55.26	132.56	209.87	0.75 (fix)	M_{33}	-17.33	53.40	124.12	0.75 (fix)
M_{30}	44.86	116.38	187.91	0.95 (fix)	M_{33}	-64.26	3.68	71.62	0.95 (fix)

† $\bar{D}$ is the posterior mean deviance $E_\theta[-2\log f(\boldsymbol{y}|\boldsymbol{\theta})]$; ‡ $\hat{D} = -2\log f(\boldsymbol{y}|\bar{\boldsymbol{\theta}})$ where $\bar{\boldsymbol{\theta}}$ is the posterior mean of $\boldsymbol{\theta}$

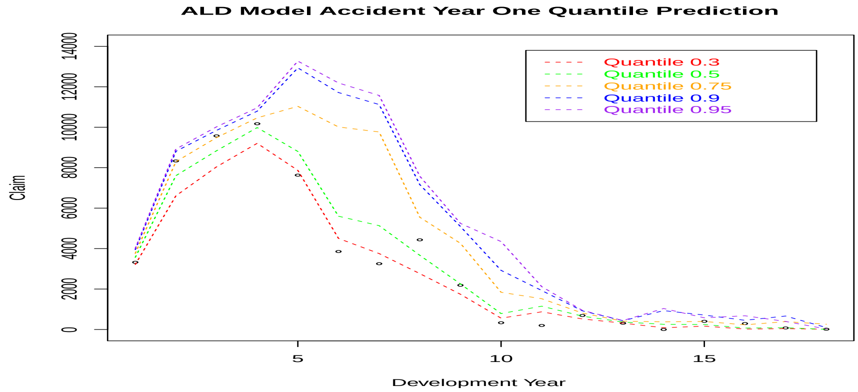
Between parametric model M_{23} and nonparametric models M_{33}, the nonparametric models provide better model performance according to DIC. These models correspond to the AL models with mean and variance functions and we study their performances for a range of fixed quantile levels $p \in \{0.3, 0.5, 0.75, 0.95\}$ as shown in Figure 3.5. This plot demonstrates the quantile-quantile plot for the fitted models at different quantile levels, indicating appropriate fits from the specified model structures for a range of different quantile levels.

FIGURE 3.5. QQ plot for nonparametric models M_{33} at different quantile levels



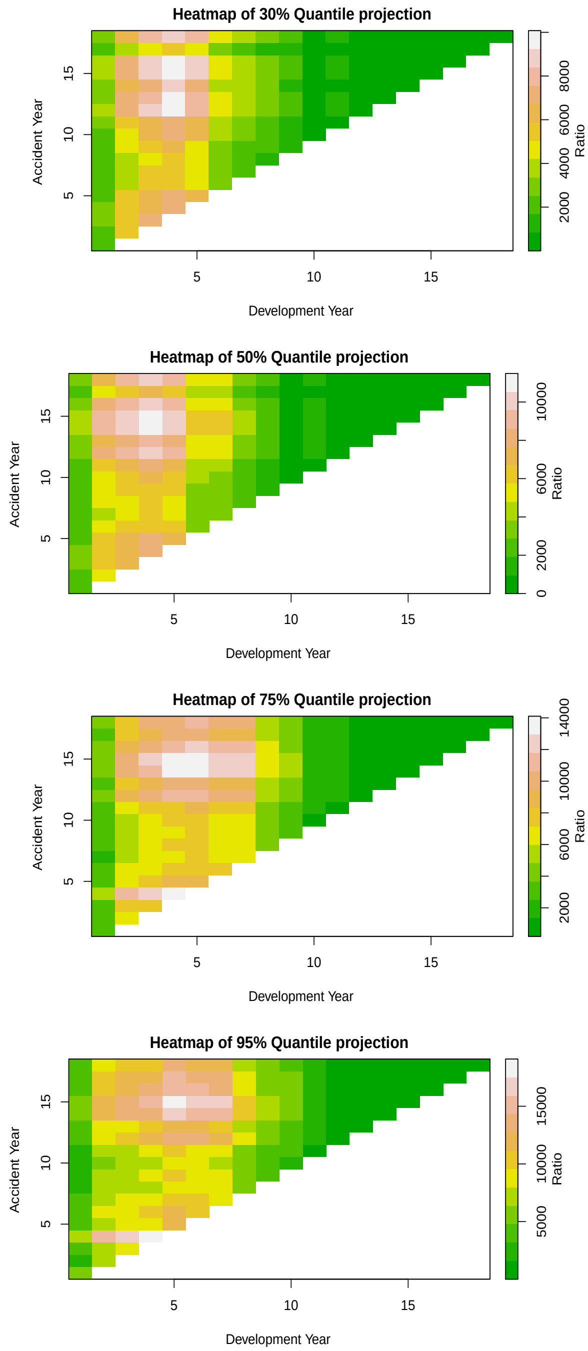
In addition, we investigate the trends of development year effects as depicted in Figure 3.6 which reports the fitted loss $\widehat{Y}_{1j} = \exp(\mu^*_{1j})$ where μ^*_{1j} is given by (28) and calculated using the conditional predictive posterior quantile function in (44) for the first accident year ($i = 1$). The quantile levels u correspond to the shape parameter p set to 0.3, 0.5, 0.75 and 0.95 respectively in AL distribution. The figure demonstrates that there is a clear requirement for a nonlinear trend in the development year covariate at all quantile levels which uniformly increases up until $j = 4$ and subsequently decreases thereafter at all quantile levels. Furthermore, the trends of fitted loss at all quantile levels agree with this observed trend.

FIGURE 3.6. Fitted loss of the first accident year across quantiles using M_{33} with AL distribution



To conclude the benchmark analysis on model structure we also present for the best model M_{33} with mean and variance functions the estimated model trends for all accident years, depicted in Figure 3.7 as five triangular heat maps. The heat maps each depict the fitted loss by accident and development years in the upper triangle at all five quantile levels, where the first row corresponds to that which was studied in Figure 3.6. All heat maps show a consistent trend across development years for all accident years and quantile levels with high levels of loss as indicated by light colours being around the fourth development year, particularly for lower accident years. With increasing quantile levels, the width of light colours for each accident year increases showing higher levels of fitted losses around the peak.

FIGURE 3.7. Fitted loss of the upper triangle across quantiles using M_{33} with AL distribution



Although nonparametric models have lower DIC values, Table 3.1 shows that parametric model M_{23} actually provides comparable model fit according to $\bar{D}$s before model complexity penalty was applied. This is because parametric models with additional shape parameters are subject to heavier model complexity penalty. However it should be noted that parametric models provide better model fit in general over a range of models and quantile levels. In addition, the parametric models have a significant advantage that they will be more readily interpretable as well as directly usable when calculating risk margins and quantile based risk measures as long as the quantile functions are in closed form, as was discussed in Section 3.4. For the mean structure corresponding to model choice $M_{2\cdot}$ under parametric model we also studied different variance structures, in order to explore the different choices of variance functions under the AL distribution.

TABLE 3.2. Parameter estimates and model fit measures for AL models with ANOVA mean and various variance functions

Models	DIC	$\bar{D}$	$\hat{D}$	MSE	p	σ^2
M_{20}	50.94	120.17	189.40	1015.71	0.80	0.02
M_{21}	-4.32	56.66	117.64	849.91	0.74	0.04
M_{22}	6.63	54.29	101.95	755.66	0.68	0.19
M_{23}	-20.81	24.91	70.63	850.10	0.75	0.17

Again, we confirm that amongst all models with AL distribution, M_{23} which incorporates both accident and development year effects for the mean and variance demonstrates the best model fit according to DIC. On the other hand, MSE favors M_{22} which adopts only development year effect for the variance. One possible reason might be that the payments made in different accident years are relatively stable compared to those across development years, and hence the development year effect dominates in the variance estimation.

3.5.2. Analysis of Quantile Regression Models: Quantile Distribution.

In this section we analyze the different model choices from the distributional perspective. This is not directly trivial to achieve, since each model has different features that must be taken into consideration in the comparison. It is clear from previous studies that one should always utilize an ANOVA-type mean function with accident and development years effect ($M_{2\cdot}$), or at a minimum incorporate a quadratic or basis function form for the development year effects such as $M_{1\cdot}$. In the case of the GB2 and AL models we will therefore consider mean structures in $M_{2\cdot}$. However, in the case of the PP model we will consider $M_{1\cdot}$, since purely from a computational perspective it will be easier to implement an efficient MCMC sampler for $M_{1\cdot}$ compared to $M_{2\cdot}$. The reason for this is due to the rejection stage in the Metropolis-Hastings acceptance probability where under the PP model the posterior constraint regions will be easier to satisfy with less model complexity. In terms of the variance functions, when working with the GB2 models, we will consider $M_{2\cdot}$ in which we do not specify variance functions as there is no variance parameter in the distribution to model the variance directly. The variance of the models are given by (19). Then in the case of the AL model we consider M_{20} as well as M_{23} and for the PP model we consider M_{10} and M_{13}.

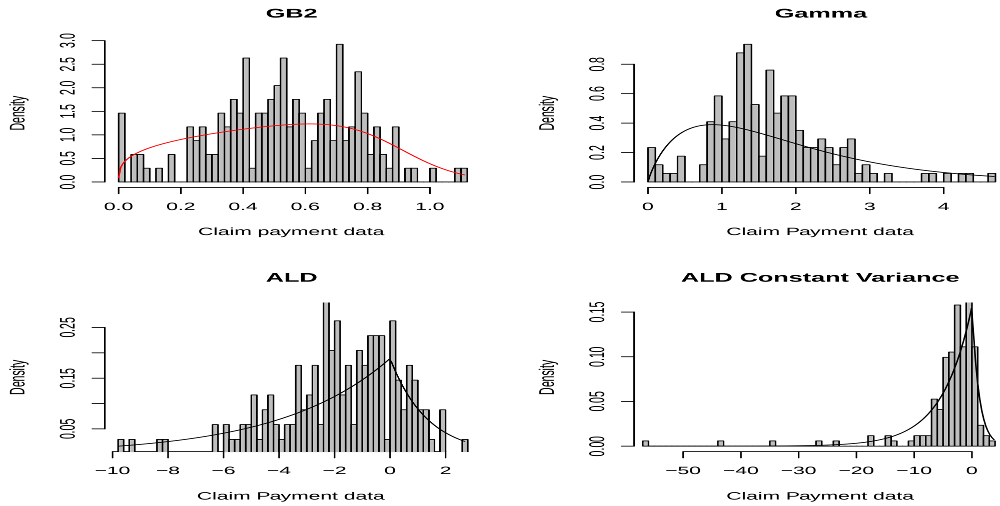
Table 3.3 reports the results split according to models with constant, unspecified and dynamic variance functions. In the case of constant or unspecified variance, the best performing model is again the AL model, followed by the GG model. Among distributions in the GB2 family with positive support, GG provides the best model fit according to DIC with model complexity penalty while GB2 model provides the best model prediction according to MSE. Comparing $\bar{D}$s without model complexity penalty, GG and GB2 provide very similar model fit. Besides, it is clear that the PP model with only the basis function regression structure for the mean, given by a quadratic polynomial for the trend in the development year covariate, and a constant variance was not sufficient to capture all the features required. We believe that this is largely due to the fact that such a model is more suitable for heavy tailed run-off in the claims development and the Israel data clearly does not display such a feature. It is therefore expected that such a heavy tailed quantile regression model will not perform as well for this data. When the variance is also modeled, the AL model is clearly significantly better than all the other models considered, again making M_{23} with AL model optimal compared to all choices. Since, the PP model is shown to be not suitable for this data, we will consider analyses going forward with only the GB2 and AL models.

TABLE 3.3. Parameter estimates and model fit measures for models with various distributions

Models	DIC	$\bar{D}$	$\hat{D}$	MSE	a	p	q	σ^2
Quantile Regression: Unspecified Variance Function								
$M_{2\cdot}$ Gamma	3064.50	3028.93	2993.36	537.82	1	1.87	∞	-
$M_{2\cdot}$ GG	2707.42	2932.97	3158.52	582.78	33.22	0.08	∞	-
$M_{2\cdot}$ GB2	3002.82	2964.60	2926.37	526.65	-7.94	1.78	0.17	-
Quantile Regression: Constant Variance Function								
M_{10} PP	3272.14	1021.71	1230.01	1132.12	-	-	-	14.15
M_{20} AL	50.94	120.17	189.40	1015.71	-	0.80	-	0.02
Quantile Regression: Non-Constant Variance Function								
M_{13} PP	1502.19	1906.49	2310.98	923.00	-	-	-	9.10
M_{23} AL	-20.81	24.91	70.63	850.10	-	0.75	-	0.17

Next, we compare the standardized residuals for the GB2, Gamma and GG models under structure in $M_{2\cdot}$ against the best fitting AL model, that is M_{23} with $\hat{p} = 0.75$. We first assess how well these models perform in sample, by looking at the following fitted model densities, versus the histograms of standardized residuals, displayed in Figure 3.8. This plot shows that $M_{2\cdot}$ with GB2 distribution and M_{23} with AL distribution and $\hat{p} = 0.75$ provide good fit to the standardized residuals whereas gamma distribution provides the worst fit.

FIGURE 3.8. Standardized residual plot



Then, for out of sample analysis we display in Figure 3.9 the median predicted total claim reserve under the GB2 and AL ($p = 0.5$) models. To compare these models for the out-of sample predictions we compare fitted losses of the four models by plotting $\widehat{Y}^{(\mathrm{p})}$ against the percentile p where $\widehat{Y}^{(\mathrm{p})}$ refers to the p-th percentile of all $\widehat{Y}_{ij} = \mu_{ij}$ in the upper triangle arranged in ascending order. We can see that the fitted losses using AL model are closest to the observed losses, GG and GB2 models provide very similar fitted losses and gamma model provides the poorest fit.

FIGURE 3.9. Percentiles of fitted losses in the upper triangle using GB2 family and AL distributions

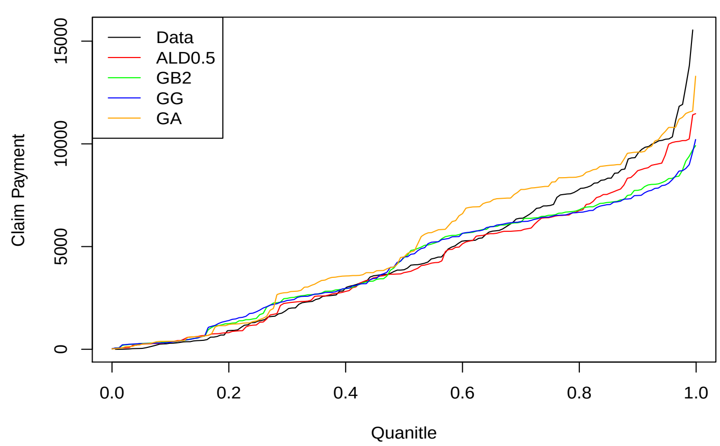
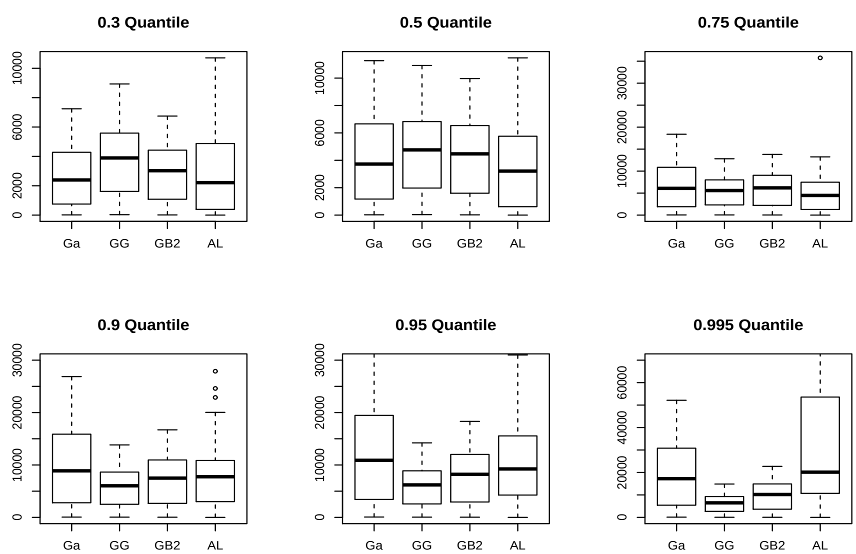


TABLE 3.4. Selected percentiles of fitted losses in the upper triangle using GB2 and AL models

Models	0.30	0.50	0.75	0.90	0.95
Observed	1,985	3,871	6,990	9,327	10,200
M_2. Gamma	2,760	4,496	8,036	9,600	10,700
M_2. GG	2,378	4,498	6,451	7,486	8,040
M_2. GB2	2,480	4,463	6,526	7,737	8,247
M_{23} AL ($p = 0.5$)	2,255	3,734	6,422	8,696	9,715

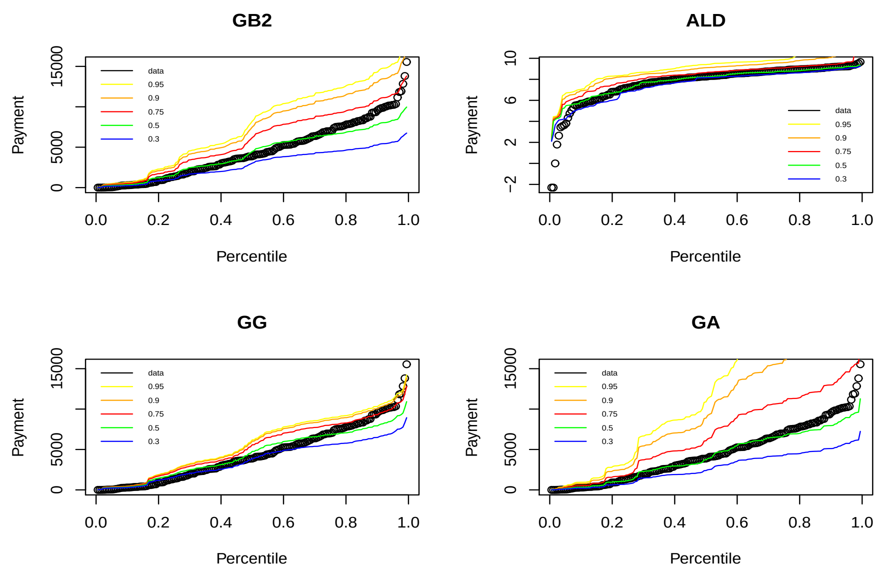
Table 3.4 reports the observed and fitted loss $\widehat{Y}^{(\mathtt{p})}$ for $\mathtt{p}$ = 0.3, 0.5, 0.75, 0.9 and 0.95 using the four models. As the model assessments show adequate model fits, we apply the models to predict losses at different quantile levels. Figure 3.10 presents boxplots of quantiles $Q_Y(u|\boldsymbol{x}_{ij})$ for losses in each cell of the upper triangle for a given quantile level u and model. Comparing across models, the boxplots for AL model have the heaviest right tails and the ranges of boxplots differ more at higher quantile level. In particular, the ranges for gamma and AL models increase much faster across quantile levels than the GG and GB2 models.

FIGURE 3.10. Boxplots of predicted quantile in the upper triangle using GB2 family and AL distributions



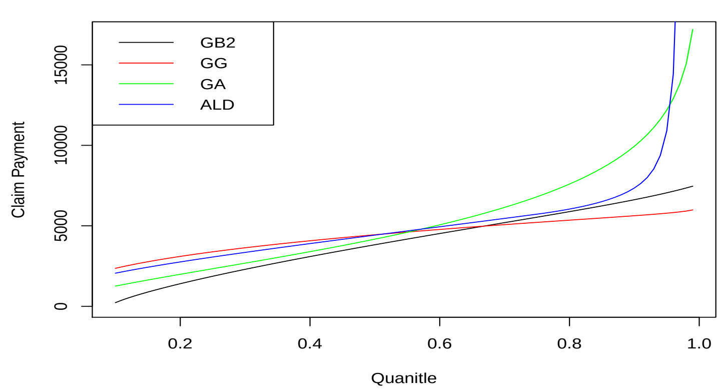
These features can also be observed in Figure 3.11 which plot quantiles $Q_Y(u|\boldsymbol{x}_{ij})$ in each boxplot in ascending order. This is similar to Figure 3.9 but the percentile of quantiles $Q_Y(u|\boldsymbol{x}_{ij})^{(\mathtt{p})}$ instead of fitted $\widehat{Y}_{ij}^{(\mathtt{p})}$ is plotted against the percentile $\mathtt{p}$. Each line in Figure 3.11 corresponds to a quantile level $u = 0.3, 0.5, 0.75, 0.9$ and 0.95. These so called empirical quantile lines are dense for GG model, sparse for gamma model and moderate for GB2 model indicating that GB2 distribution provides quantile estimates which can reasonably cover the observed losses across percentile p when the quantile level u gradually increases. We also remark that the empirical quantiles for AL model in the log scale are convex rather than concave and are more dense because of the log transformation.

FIGURE 3.11. Percentiles of predicted quantiles in the upper triangle using GB2 and AL models



Then Figure 3.12 plots the quantile functions $Q_Y(u|\boldsymbol{x}_o)$ across quantile levels $u \in (0, 1)$ using (22) for gamma, GG and GB2 in the GB2 family of distributions and $\exp(Q_{Y^*}(u|\boldsymbol{x}_o))$ in (8) where $Q_{\epsilon^*}(u) = F_{z^*}^{-1}(u)$ is given by (11) for AL distribution. Note that the mean μ in $Q_Y(u|\boldsymbol{x}_o)$ or μ^* in $\exp(Q_{Y^*}(u|\boldsymbol{x}_o))$ is given by the average of $\exp(\mu^*_{ij})$ or μ^*_{ij} over risk cells in the upper triangle. Again AL distribution has the heaviest right tail because of the log transformation.

FIGURE 3.12. Quantile functions using GB2 family and AL distributions



We further adopt these models to calculate the outstanding reserves (OR) as reported in Table 3.5 using the conditional predictive posterior approach in (47) where the conditional total reserve posterior quantile function is adopted for the case of light tailed run-off in the claim process because the claim distribution was shown to be light tailed in the previous analyses. Under the Solvency II framework, insurers will have to establish technical provisions to cover future claims expected from policyholders. Insurers must also have available financial resources sufficient to cover both a minimum capital requirement and a SCR. The SCR is based on a VaR measure calibrated to a 99.5 percent confidence level over a one-year time horizon. Results in Table 3.5 show that the OR projection increases gradually up to 95 percentile quantile levels but increases dramatically at 99.5 percentile.

TABLE 3.5. Outstanding reserves at different quantile levels using GB2 family and AL distributions

Models	0.30	0.50	0.75	0.90	0.95	0.995
$M_{2\cdot}$ Gamma	127,816	198,907	324,515	474,073	581,302	920,142
$M_{2\cdot}$ GG	203,207	248,409	291,457	314,482	323,346	337,658
$M_{2\cdot}$ GB2	152,315	225,017	311,625	377,154	413,525	512,731
M_{23} AL	145,031	176,926	314,454	435,402	462,980	560,430

3.6. Risk Margin: Australian Case Study

In general the guidance on calculation of risk margin by regulators leaves flexibility in the practical modelling approach adopted by practitioners. There are a few popular approaches considered in practice, some of which involve a degree of expert opinion. In this section we aim to consider only approaches based on statistical models and in particular percentile and quantile based methods. In this context the standard practice is to consider the reserve estimate and then try to quantify the uncertainty associated with the reserve estimator. This uncertainty is typically measured via a standard error, which is utilized to adjust the reserve. Traditionally, if a loss distribution produces an estimator for the reserve which admits a normal distribution (approximately under a central limit theorem result), then setting the risk margin to equal the sample estimator for the reserve plus 0.675 times the sample estimators standard deviation would result in risk margins calibrated to approximately the 75th percentile. Note, whilst the total loss distribution may not have finite second moment if a heavy tailed run-off is present, the variance of the sample estimator for the distribution of the reserve will always be well defined. It should be noted that this method suffers from drawbacks as there is both an influential judgment in determining the appropriate multiple, especially when the normality assumption is not present due to sample estimators distribution being skewed.

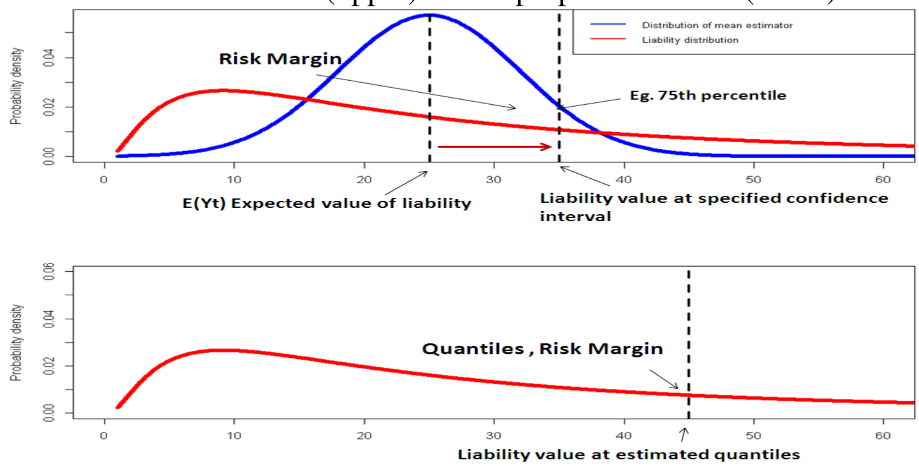
Alternatively, one may utilize the quantile regression model obtained for the total loss distribution. There are two basic ways this may be achieved, for instance one could take instead of a mean reserve, a quantile based reserve. This could be via a risk measure such as VaR which represents a tail quantile of the total loss distribution at say 99.95%, in which case one may judge that a conservative measure of reserve is obtained from such a tail measure and so no additional risk margin is required. This is standard in banking regulations such as Basel II/III and being considered in insurance regulations.

Alternatively, one may take a central measure as the reserve such as the median of the total loss distribution and make a risk margin adjustment based on the tail quantile of the total loss distribution at say 75% (as is considered in practice).

Thirdly, if the traditionally utilized estimate of reserve based on the mean of the loss distribution is considered, then two scenarios may arise if one uses the risk margin adjustment based on the tail quantile of the total loss distribution at say 75%. In this case the estimated mean reserve could be below the desired risk margin quantile level of the total loss distribution, in which case it may be reasonable to make no further adjustment if the risk margin is already at a tail quantile such as 75%. Alternatively, if the estimated mean reserve is below the desired risk margin quantile level of the total loss distribution, then the difference would be the resulting risk margin.

In this section, we are going to extend the best model, model M_{23} with AL distribution, in the previous sections to model risk margin statistically. To achieve this, we generalise the AL distribution to model the shape parameter p via the following regression $p_i = \phi_0 + \phi_i$ where ϕ_0 is the intercept and ϕ_i denotes accident year effect. Accident year effect is chosen because risk capital allocation is by accident years. It is worth noting an important assumption which are stated as underlying this method: actual outstanding claim payments are assumed to be uncorrelated between accident years. Therefore, the estimated shape parameter p, which presents quantile in AL distribution, and also infers risk margin in the percentile method, is an applicable risk margin estimate for outstanding claims payments. The difference between our proposed method and the traditional method is also demonstrated in Figure 3.13.

FIGURE 3.13. Traditional method (upper) versus proposed method (lower)



The data that we used to demonstrate our model is the amount of payments for all the compulsory third party (CTP) policies in Queensland (QLD) as of June 2008. CTP insurance policy covers risk that would be referred to as Auto Bodily Injury in the U.S. and Motor Bodily Injury in the U.K.. The data are in the units of millions summarized by accident and development quarters covering periods from December 2002 to June 2008. It contains 276 observations over 23 accident quarters. In order to remove the influence of inflation for reserving purposes, we utilize the average weekly earning index from the Australian Bureau of Statistics (ABS) to inflate all the values to December 2008 dollars. Hence, the data used in this analysis represents the inflated cumulative payment for QLD CTP portfolio as reported in Figure 3.19 in appendix I.

FIGURE 3.14. Observed variance of QLD CTP payment data by accident year

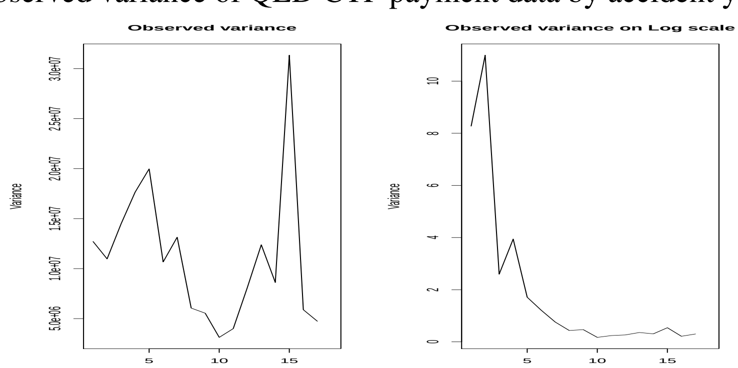
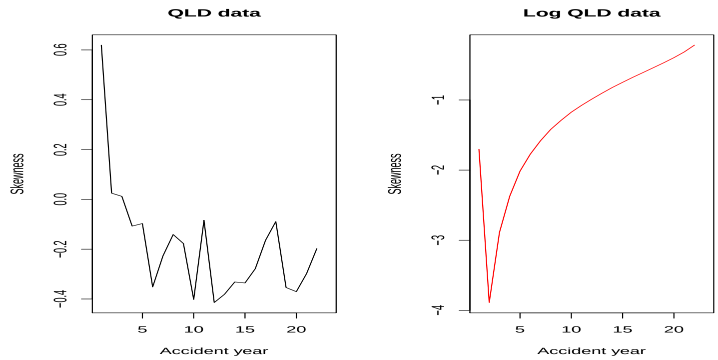


FIGURE 3.15. Observed skewness of QLD CTP payment data by accident year



To review features of the data, Figure 3.14 plots the observed variance across accident year on original and log scale. It shows that the variance fluctuates a lot across accident year on the original scale but displays a sharp drop on the log scale. Figure 3.15 shows that the skewness are mostly negative on the original and log scales. The overall skewness of the data is 0.61 and that on a log scale is -1.08. Trend of skewness reveals a sharp drop at the start and then it fluctuates across accident years for data on the original scale but increases monotony for data on the log scale. These changes confirm the necessity of adopting dynamic variance and skewness in modelling the data.

Among choices of distributions, the AL distribution allow flexibility in modelling variance and skewness through modelling directly the scale and shape parameters σ^2 and p respectively. Furthermore, in the context of nonparametric regression using AL as a proxy distribution for model implementation, p indicates the quantile level of a model which corresponds to risk margin in loss reserving. In the analysis of QLD CTP data, we adopt the ANOVA type model (M_{23}) for the mean and variance as it has been shown to provide the best model performance. We further propose modelling the risk margin p as a linear function of accident year. One reason is that as accident year increases, there are more uncertainty involved in estimating the reserves; hence it is an important factor in risk margin estimation. This model is called $M_{23'}$ in the Appendix.

Then $M_{23'}$ with dynamic variance and skewness is compared to two models, M_{20} with constant variance and skewness and M_{23} with just dynamic variance in Table 3.6. Although M_{20} outperform $M_{23'}$ according to DIC, $M_{23'}$ provides the best model fit according to $\bar{D}$ which measures model fit alone, discounting model complexity penalty. As our aim is to provide the most accurate risk margin estimates, we adopt $M_{23'}$ in the subsequent risk margin analysis. From a modelling perspective, it reconciles with our risk margin estimation approach.

TABLE 3.6. Parameter estimates and model fit measures for ANOVA models using QLD CTP payment data

Models	DIC	$\bar{D}$	$\hat{D}$	$E(Y)$	$Var(Y)$	$S(Y)$
M_{20} Constant variance & skewness	-322.55	-215.65	-108.75	4.33	0.008	-0.28
M_{23} Dynamic variance	-311.36	-197.71	-84.06	7.67	0.22	-0.57
$M_{23'}$ Dynamic variance & skewness	-255.03	-229.46	-203.90	4.77	0.10	-0.18

Figure 3.16 demonstrates how the estimated risk margin $\hat{p}_i$ changes across accident years, superimposed with its creditable interval. Figure 3.17 displays the corresponding changes in estimated variance and skewness using the variance and skewness equations in (12) and (13) respectively. The risk margin $\hat{p}$ starts at 0.895 at accident year 1 when the variance is quite high. Afterwards, it decreases gradually to 0.439 in accident year 8 when the variance is much smaller. From accident year 17 onwards, the risk margin increases again when the variance is large and there are more development years ahead. In actuarial practice, the calculation of the risk margin is often not based on a sound model but various simplified methods are used. This approach enables us to calculate a risk margin for non-life insurance run-off liabilities in a mathematically consistent way, and provides reasonable risk margin estimates.

FIGURE 3.16. Change of p across accident year using $M_{23'}$ for risk margin analysis

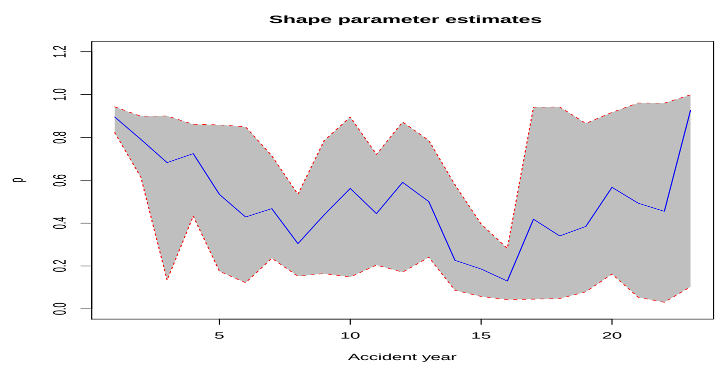
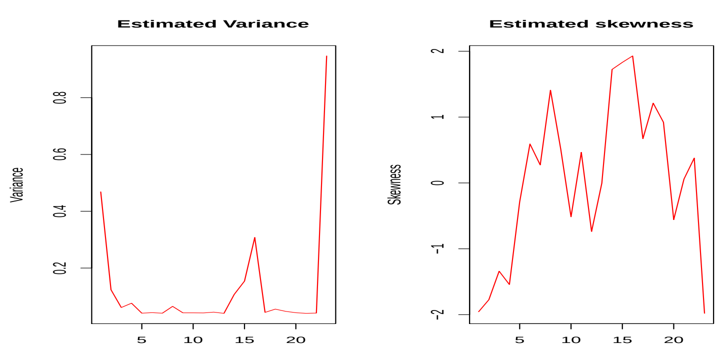


FIGURE 3.17. Estimated variance and skewness in $M_{23'}$ for risk margin analysis



3.7. Conclusion

We have applied the quantile regression model to estimate loss reserve and risk margin. Quantile regression reveals relationships between responses at the upper or lower quantiles, which is of significant interest in estimating risk margin and VaR in insurance and finance applications. Compared to mean regression, it is more robust to heavy tailed data. We compare the performance of parametric and non-parametric quantile regression. In the parametric framework, we built five models, namely AL, PP, GB2, GG and gamma. The AL model provides the best fit. We also investigate three different regression structures, namely ANCOVA, ANOVA and Poisson-Tweedie regression. The ANOVA model performs the best in our empirical data study.

Furthermore, we adopt the best performed model, which is the AL model with ANOVA mean and variance functions, to estimate risk margin. The generalized AL model with a dynamic shape parameter p provides us a mathematically consistent way of estimating risk margin. Overall, the results of our studies indicate that this new risk margins framework offers considerable potential benefits for reserving purpose. However, the drawback is that quantile functions may cross over particularly at extreme quantiles when data are rare. Extreme quantile may not be estimated precisely. Although there is no simple solution to this problem yet, we believe it is important to be aware of this limitation when using this framework.

FIGURE 3.18. Israel payment data

	Development Quarter																	
Accident Quarter	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	3,323	8,332	9,572	10,172	7,631	3,855	3,252	4,433	2,188	333	199	692	311	0.01	405	293	76	14
2	3,785	10,342	8,330	7,849	2,839	3,577	1,404	1,721	1,065	156	35	259	250	420	6	1	0.01	
3	4,677	9,989	8,746	10,228	8,572	5,787	3,855	1,445	1,612	626	1,172	589	438	473	370	31		
4	5,288	8,089	12,839	11,829	7,560	6,383	4,118	3,016	1,575	1,985	2,645	266	38	45	115			
5	2,294	9,869	10,242	13,808	8,775	5,419	2,424	1,597	4,149	1,296	917	295	428	359				
6	3,600	7,514	8,247	9,327	8,584	4,245	4,096	3,216	2,014	593	1,188	691	368					
7	3,642	7,394	9,838	9,733	6,377	4,884	11,920	4,188	4,492	1,760	944	921						
8	2,463	5,033	6,980	7,722	6,702	7,834	5,579	3,622	1,300	3,069	1,370							
9	2,267	5,959	6,175	7,051	8,102	6,339	6,978	4,396	3,107	903								
10	2,009	3,700	5,298	6,885	6,477	7,570	5,855	5,751	3,871									
11	1,860	5,282	3,640	7,538	5,157	5,766	6,862	2,572										
12	2,331	3,517	5,310	6,066	10,149	9,265	5,262											
13	2,314	4,487	4,112	7,000	11,163	10,057												
14	2,607	3,952	8,228	7,895	9,317													
15	2,595	5,403	6,579	15,546														
16	3,155	4,974	7,961															
17	2,626	5,704																
18	2,827																	

FIGURE 3.19. QLD CTP payment data

	Development Quarter																							
Accident Quarter	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**Exposure**
Dec-02	0.1	0.6	1.3	2.1	3.0	4.3	5.8	5.6	4.9	6.7	6.1	14.3	13.0	9.8	7.9	8.9	8.0	4.0	9.4	3.4	3.3	2.2	5.4	2.6
Mar-03	0.1	0.6	0.9	0.9	1.3	1.8	3.0	2.6	2.7	5.9	6.1	7.4	7.9	8.5	9.1	9.3	5.4	6.6	8.4	5.3	1.4	2.5		2.6
Jun-03	0.1	0.6	0.7	0.8	1.2	1.7	2.6	3.0	4.3	6.0	6.9	5.8	7.1	10.2	11.0	5.5	8.1	6.2	4.8	1.3	7.3			2.6
Sep-03	0.1	0.6	0.8	1.0	1.3	2.0	2.1	3.6	4.5	8.0	5.8	5.4	11.0	9.6	6.5	10.5	7.4	6.2	5.4	4.0				2.6
Dec-03	0.0	0.6	0.9	1.1	1.0	1.1	2.5	3.8	5.2	4.5	7.4	6.0	12.1	5.3	9.7	8.0	6.7	4.4	8.2					2.7
Mar-04	0.1	0.5	1.0	0.9	0.9	1.7	2.4	4.2	5.0	8.1	7.5	8.7	7.9	9.9	8.6	11.8	10.0	5.6						2.7
Jun-04	0.1	0.5	0.9	0.7	1.0	1.8	2.6	4.1	5.8	5.6	7.2	6.2	10.0	6.5	11.5	6.8	5.4							2.7
Sep-04	0.1	0.6	0.7	0.9	1.1	2.1	3.2	4.8	5.8	7.5	8.0	14.2	9.4	15.0	8.4	11.7								2.8
Dec-04	0.1	0.5	0.8	0.9	1.1	1.8	4.1	5.9	6.5	5.9	12.5	9.2	13.9	6.2	10.0									2.8
Mar-05	0.0	0.5	1.0	0.8	1.1	1.5	4.1	6.7	5.5	10.7	7.9	8.2	8.8	8.7										2.8
Jun-05	0.1	0.7	0.8	0.9	1.3	2.3	5.3	4.4	7.4	8.3	9.1	7.3	15.2											2.9
Sep-05	0.1	0.6	1.1	0.9	2.3	2.9	3.8	8.2	8.9	9.6	6.9	8.6												2.9
Dec-05	0.1	0.7	1.0	1.2	1.5	3.0	5.7	7.2	8.7	7.8	9.8													2.9
Mar-06	0.0	0.4	0.6	0.8	0.9	3.2	4.7	7.2	6.8	7.0														3.0
Jun-06	0.1	0.6	0.8	0.7	1.9	4.4	7.5	6.5	7.7															3.0
Sep-06	0.0	0.5	0.7	0.8	1.9	6.2	6.9	7.7																3.0
Dec-06	0.1	0.5	1.0	1.5	2.0	4.4	7.6																	3.1
Mar-07	0.0	0.7	1.0	1.0	1.6	5.6																		3.1
Jun-07	0.1	0.6	0.8	0.9	2.0																			3.2
Sep-07	0.1	0.7	0.9	1.1																				3.2
Dec-07	0.1	0.5	0.7																					3.2
Mar-08	0.1	0.6																						3.3
Jun-08	0.1																							3.3

APPENDIX II The following table shows the model structures considered for each regression analysis.

Model Index	Model Location Structure	Model Scale Structure	Distribution Types	Model Description
M_{21}	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	$\sigma_{ij} = \beta_0 + \beta_{1i}$	AL	**Location:** Fully parameterized model with individual trend components in accident and development years. **Scale:** heteroskedasticity in accident years with common variance over development years scale parameter.
M_{22}	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	$\sigma_{ij} = \beta_0 + \beta_{2j}$	AL	**Location:** Fully parameterized model with individual trend components in accident and development years. **Scale:** heteroskedasticity in development years with common variance over accident years scale parameter.
M_{23}	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	$\sigma_{ij} = \beta_0 + \beta_{1i} + \beta_{2j}$	AL	**Location:** Fully parameterized model with individual trend components in accident and development years. **Scale:** heteroskedasticity in development and accident years scale parameter.
$M_{23'}$	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	$\sigma_{ij} = \beta_0 + \beta_{1i} + \beta_{2j}$ $\quad p = \phi_0 + \phi_{1i}$	AL	**Location:** Fully parameterized model with individual trend components in accident and development years. **Scale:** homoskedasticity in scale parameter and shape parameter p (quantile level) has trend in the accident years (common across all development years).
M_{30}	$\mu^*_{ij} = \alpha_{0,u} + \alpha_{1i,u} + \alpha_{2j,u}$	$\sigma_{ij} = \sigma$	AL as proxy	**Location:** Nonparameterized model with individual trend components in accident and development years. **Scale:** not defined in the model.

TABLE 3.8. Model Structures in the Quantile Regressions.

Model Index	Model Location Structure	Model Scale Structure	Distribution Types	Model Description
M_{00}	$\mu^*_{ij} = \alpha_0 + \alpha_1 \times i + \alpha_2 \times j$	$\sigma_{ij} = \sigma$	AL	**Location:** Simple Additive Model (parsimonious) common trend in accident years and development years. **Scale:** homoskedasticity in development years scale parameter (common across accident years).
M_{10}	$\mu^*_{ij} = \alpha_0 + \alpha_1^S F_1(j) + \alpha_2^C F_2(j)$	$\sigma_{ij} = \sigma$	AL	**Location:** Basis function regression model with trend component for development years given by Level, Slope and Curvature components (common across accident years). **Scale:** homoskedasticity in development years scale parameter (common across accident years).
M_{20}	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	$\sigma_{ij} = \sigma$	AL, PP	**Location:** Fully parameterized model with individual trend components in accident and development years. **Scale:** homoskedasticity in development years scale parameter (common across accident years).
$M_{2\cdot}$	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	Eqn 19.	GB2	**Location:** Fully parameterized model with individual trend components in accident and development years.

TABLE 3.7. Model Structures in the Quantile Regressions. Note: basis function choices $F_1(j) = \left(\frac{1-e^{-\lambda \times j}}{\lambda \times j}\right)$, $F_2(j) = \left(\frac{1-e^{-\lambda \times j}}{\lambda \times j} - e^{-\lambda \times j}\right)$.

APPENDIX III

Model Selection

To compare between different models, we utilise the deviance information criteria (DIC) and the mean sum of squared error MSE. The former is a Bayesian analogue of Akaike's Information Criterion (AIC) which is commonly used in Bayesian analysis. It consists of a measure of model fit which is the posterior mean deviance $\bar{D} = E_\theta[-2\log f(\boldsymbol{y}|\boldsymbol{\theta})]$, and a measure of model complexity which is an estimate of the effective number of parameters $\hat{p}$ given by the difference between posterior mean deviance $\bar{D}$ and $\hat{D} = -2\log f(\boldsymbol{y}|\bar{\boldsymbol{\theta}})$.

The DIC is given by

$$DIC = -\frac{4}{K}\sum_{k=1}^{K}\sum_{i=1}^{I}\sum_{j=1}^{I-i+1}\ln\left[f(y_{ij}|\boldsymbol{\theta}^{(k)})\right] + 2\sum_{i=1}^{I}\sum_{j=1}^{I-i+1}\ln\left[f(y_{ij}|\overline{\boldsymbol{\theta}})\right] \qquad (50)$$

where $\boldsymbol{\theta}^{(k)}$ denotes the vector of parameter estimates in the k-th iteration of the posterior sample of size K, $\bar{\boldsymbol{\theta}}$ denotes the posterior mean of $\boldsymbol{\theta}^{(k)}$ and $f(y_{ij}|\boldsymbol{\theta})$ represents the densities in (3.2), (3.3) and (16) respectively for AL, PP and GB2 distributions where μ_{ij}^* and σ_{ij}^2 are given by the models in Section 2.3. The advantage of DIC over other criteria in Bayesian model selection is that the DIC is easily calculated from the samples generated by an MCMC simulation. Claeskens and Hjort (2008) show that the DIC is large-sample equivalent to the natural model-robust version of the AIC.

While the DIC measures the model fit and penalizes model complexity, the MSE defined as

$$MSE = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{I-i+1}(y_{ij} - \mu_{ij})^2$$

assesses the accuracy of model prediction by comparing the observed y_{ij} with fitted μ_{ij} for losses in the upper triangle. Obviously, models with small MSE indicate close agreement and hence provide good models predicting losses in the lower triangle.

CHAPTER 4

A Copula Based Paid-Incurred Claims Models

The Bayesian reserving model based on payment data has been extended to quantile functions to derive risk margin, as discussed in Chapter 3. This Chapter extends the payment based models to combine claims payments and incurred losses information into a coherent reserving methodology, using a Data-Augmented mixture Copula Paid-Incurred claims model.

4.1. Background

As discussed in Merz and Wuthrich. (2010) the main task of reserving actuaries is to predict ultimate loss ratios and outstanding loss liabilities. In general such predictions are based on past information that comes from a variety of sources. Under a credibility based framework, the weighting of different data sources and their relative contribution to the estimated reserve is difficult to determine. Therefore, it is important to consider developing a unified prediction framework for the outstanding loss liabilities, known as the paid-incurred-claims (PIC) class of models. However, to date only simple dependence structures have been considered, with three parameters for the correlations which were not incorporated into the formal Bayesian estimation approach, and instead fixed deterministically *a priori*. There are two technical difficulties in extending the current restrictive assumptions within a Bayesian framework. The first is being able to generate the positive definite matrices; the second is evaluating the joint likelihood of the mixture copula defined over the observed payments and incurred losses in each accident year row of the reserving matrix. Our article significantly extends the dependence structure of current PIC models by solving these two problems. The first problem is resolved through utilisation of a class of matrix-variate Inverse-Wishart priors coupled with an adaptive Markov chain sampler that restricts the proposed Markov chain states to remain on the manifold of such matrices. The second problem is solved by using a data augmentation strategy which treats the unobserved parts of the loss triangle as missing data so that one can perform evaluation of the copula based likelihood required for inference on the model parameters.

In order to ensure the financial security of an insurance company, it is important to predict future claims liabilities and obtain the corresponding prediction intervals which take into account parameter uncertainty. The PIC model is a claims reserving method which statistically combines information about claims payments and incurred losses. It allows actuaries to best utilise the available information for loss reserves. The Munich chain ladder method introduced by Quarg and Mack (2004) is one of the first claims reserving approaches in the actuarial literature to unify outstanding loss liability prediction based on both sources of information. This method aims to reduce the gap between the two chain ladder predictions that are based on claims payments and incurred losses data, respectively. It is achieved by adjusting the chain ladder factors with paid-incurred ratios to reduce the gap between the two predictions. The main drawback with the Munich chain ladder method is that it involves several parameter estimates whose precisions are difficult to quantify within a stochastic model framework.

Merz and Wuthrich. (2010) recently introduced a log-normal PIC chain model and used Bayesian methods to estimate the missing (future) part of the claims reserving triangles based on both payment and loss incurred information. Its major advantage is that the full predictive distribution of the outstanding loss liabilities can be quantified. One important limitation of the model of Merz and Wuthrich. (2010) is that it does not develop the dependence properties of the PIC model that will be applicable to loss reserving data observed in practice. Our thesis extends the proposed Bayesian PIC models of Merz and Wuthrich. (2010) to capture additional dependence structures.

4.1.1. Brief Background. Dependence within payment data, within incurred loss data, and even between payment and incurred loss data commonly exists due to the nature of the loss process. Payment and incurred loss ratios in the previous development period are likely to impact that of the next development period. Hence, correlation between development periods is practically appealing in claims reserving practice. Moreover, incurred loss is essentially payment data plus case estimates which are projections foreseen by case managers to estimate the remaining payments. Correlation between payment and incurred loss data is also found. Happ and Wuthrich (2011) propose a fixed covariance structure to describe the correlation between payment and incurred loss, assuming that the correlations between different development periods are identical. In reality, correlations differ across development periods for various reasons, such as different stages of the life cycle for a claim and internal policy changes. In order to fully incorporate the actual correlations, we introduce a block covariance structure to allow for the variation between different development periods within payment and incurred losses. We also develop a second class of hierarchical mixture of copulas models.

The estimation challenge involves constructing and sampling from the resulting Bayesian models for PIC with flexible dependence structures. To specify the model, we vectorize the triangular random structures for payments and incurred loss and, applying appropriate permutations, we then assume a copula dependence structure on the vectorized data. We use a Gaussian copula with an unknown correlation matrix, which is restricted to be block diagonal for parsimony, or a mixture Archimedian copulas across development periods.

We estimate the Bayesian models by MCMC methods, using data augmentation to generate missing data values in the loss triangle and use an adaptive Metropolis algorithm to generate the unknown parameters. Bayesian simulation methodology is used to carry out inference on all aspects of the models considered and to obtain predictive distributions for reserves.

4.1.2. Contributions. We design a novel class of PIC models and illustrate it with two examples. The first involves a mixture of Clayton and Gumbel copulas for upper and lower tail dependence features in the development years for payments and incurred losses. The second example involves a Gaussian copula model in which the covariance structure is a telescoping block diagonal form representation which captures dependence between development lag years in the payments and incurred losses. By a telescoping block diagonal matrix we mean one in which the main diagonal is comprised of sub-blocks for which each incremental sub-block contains one less row and column compared to the previous. In constructing these models we consider hierarchical Bayesian models with hyperparameters on the priors for development factors and specially developed matrix-variate priors on the covariance structures which preserves the conjugacy properties of the independence models developed in Merz and Wüthrich (2010) and Merz and Wüthrich (2010).

For the independent and Gaussian copula based PIC models we develop a class of conjugate posterior models that can be efficiently estimated via an MCMC sampler known as a block Gibbs sampler. However, the extension to general copula dependence structures requires non-conjugate priors, making it necessary to develop adaptive MCMC algorithms. Adaptive sampling uses previous iterates of the Markov chain to form more efficient Metropolis proposals for the parameters, this class of MCMC algorithm has received growing attention in the statistics literature since it was recently developed and is now recognized as an important tool for Bayesian inference. There is an increasing interest in utilizing adaptive MCMC to facilitate more efficient sampling (Andrieu and Thoms. (2008), Atchadé and Rosenthal (2005)). The adaptive techniques that we adopt in this chapter fall within the general framework of adaptive Metropolis, and employ the optimal scale factors (Roberts and Rosenthal. (2009)) from the Single Component Adaptive Metropolis (SCAM) algorithm (Haario et al. (2005b)). There have been some initial utilisations of adaptive MCMC specifically in financial modelling such as Peters et al. (2011a) and the references therein. In addition the adaption strategies we consider in this paper involve extensions of Euclidean space Adaptive Metropolis to the space of positive definite matrices, creating a class of matrix variate Markov chain adaptive proposals.

In the mixture copula based PIC models, we design data augmentation strategies which are a class of auxiliary variable methods. We modify these approaches to the PIC copula based models in order to circumvent the challenge of intractable likelihood evaluations which arise form the structure of the PIC reserving triangle. In particular we argue that the tail dependence features of the model should be consistent accross all development years for both payment and incurred loss data.

This poses an evaluation challenge for the likelihood as it involves evaluation of marginal likelihood quantities given the observed data in accident year i, given by payment and incurred losses. The integral required when utilising mixture copula structures over the accident years is intractable, therefore we introduce auxiliary variables into the Bayesian model in a data-augmentation structure to overcome this dificulty.

4.2. Review of the Merz-Wuethrich Independence Copula Paid-Incurred Claims Model

This section introduces the PIC model which involves two sources of information. The first is the claims payment data, which involves payments made for reported claims. The second source of data incorporated into the statistical estimation are the incurred losses corresponding to the reported claim amounts. The differences between the incurred losses and the claim payments are known as the case estimates for the reported claims which should be equal once a claim is settled. This imposes a set of constraints on any statistical model developed to incorporate each of these sources of data into the parameter estimation. We use the constraints proposed in Merz and Wuthrich. (2010) which are used to specify a model based on a claims triangle constructed from vertical columns corresponding to development years of claims and rows corresponding to accident years. This structure for the observed data is summarized in triangular form which is utlised for both the claims payments and the incurred losses, including constraint on zero case estimates at development period J as presented in Figure 4.1.

Without loss of generality, we assume an equivalent number J of accident years and development years. Furthermore, we assume that all claims are settled after the J-th development year. Let $P_{i,j}$ be the cumulative claims payments in accident year i after j development periods and $I_{i,j}$ the corresponding incurred losses. Moreover, for the ultimate loss we assume the constraint discussed on the case estimates corresponds to the observation that predicted claims should satisfy $P_{i,J} = I_{i,J}$ with probability 1, which means that ultimately (at time J) the claims reach the same value and therefore satisfy the required constraint.

We define (i) $P_{0:J,0:j} = \{P_{k,l} : 0 \leq k \leq J, 0 \leq l \leq j\}$. (ii) Let A and B be square matrices. Then $diag(A, B)$ is the diagonal matrix, with the diagonal elements of A appearing topmost, then the diagonal elements of B. Let the matrices A and B be as in (ii). Then the direct sum of A and B, written as $A \oplus B$ is the block diagonal matrix with A in the top left corner and B in the bottom right corner. It is clear that the definitions in (ii) and (iii) can be iterated. That is $diag(A, B, C) = diag(diag(A, B), C)$ and $A \oplus B \oplus C = (A \oplus B) \oplus C$. (iv) Define the $d \times d$ diagonal square identity matrix according to $\mathbb{I}_d$. (v) Define the indicator of an event by the dirac-delta function δ_i. (vi) Define the vectorization operator on a $p \times n$ matrix A, denoted by $Vec(A)$, as the stacking of the columns to create a vector.

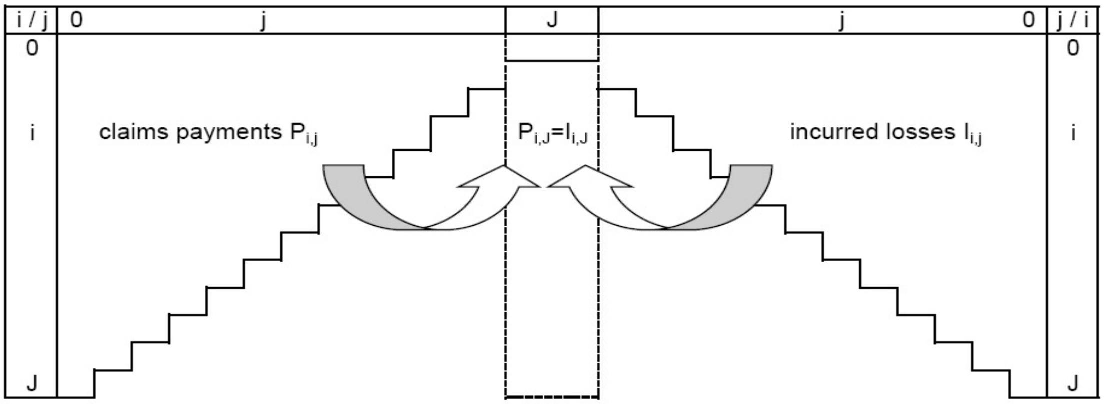


FIGURE 4.1. Claims triangle for payment data and incurred data (source Merz and Wuthrich. (2010)).

As in Merz and Wuthrich. (2010), we consider a Log-Normal PIC model as this facilitates comparison between existing results and the results we derive based on different dependence frameworks in extensions to this model.

We now introduce the PIC model and the statistical assumptions for the independent case, followed by remarks on the resulting marginal posterior models for the payment and incurred losses.

Model Assumptions 4.2.1 (Independent PIC Log-Normal (*Model I*))**.** *The model assumptions for the independent model of Merz and Wuthrich. (2010) are:*

- *The cumulative payments $P_{i,j}$ are given by the forward recursion*

$$P_{i,0} = \exp\left(\xi_{i,0}\right) \quad \text{and} \quad \frac{P_{i,j}}{P_{i,j-1}} = \exp\left(\xi_{i,j}\right) \quad \text{for } j = 1, \ldots, J$$

- *The incurred losses $I_{i,j}$ are given by the backward recursion*

$$I_{i,J} = P_{i,J} \quad \text{and} \quad \frac{I_{i,j-1}}{I_{i,j}} = \exp\left(-\zeta_{i,j-1}\right).$$

- *The random vector $(\xi_{0,0}, \ldots, \xi_{J,J}, \zeta_{0,0}, \ldots, \zeta_{J,J-1})$ has independent components with*

$$\begin{aligned} \xi_{i,j} &\sim N\left(\Phi_j\ ,\ \sigma_j^2\right) && \textit{for } i \in \{0, \ldots, J\} \textit{ and } j \in \{0, \ldots, J\}, \\ \zeta_{k,l} &\sim N\left(\Psi_l\ ,\ \tau_l^2\right) && \textit{for } k \in \{0, \ldots, J\} \textit{ and } l \in \{0, \ldots, J-1\}; \end{aligned}$$

- *The parameter vector for the model is $\boldsymbol{\Theta} = (\Phi_0, \ldots, \Phi_J, \Psi_0, \ldots, \Psi_{J-1}, \sigma_0, \ldots, \sigma_J, \tau_0, \ldots, \tau_{J-1})$. It is assumed that the components of $\boldsymbol{\Theta}$ are independent apriori. The prior density for $\boldsymbol{\Theta}$ has independent components, with σ_j, τ_j both positive for all j.*

- *It follows that*

$$\log\left(\frac{P_{i,j}}{P_{i,j-1}}\right) \sim N\left(\Phi_j\ ,\ \sigma_j^2\right) \quad \text{and} \quad \log\left(\frac{I_{i,j}}{I_{i,j+1}}\right) \sim N\left(-\Psi_l\ ,\ \tau_l^2\right) \tag{4.2.1}$$

Let $\{\mathbf{P}, \mathbf{I}\} = \{P_{i,j}, I_{k,l};\ 0 \leq i, j, k, \leq J,\ 0 \leq l \leq J-1\}$. Then, based on Model Assumptions 2.1 and the observed matrices P and I, the likelihood for $\mathbf{\Theta}$ is given by three components, see derivation in Merz and Wuthrich. (2010, Section 3.3, Equation 3.5). The first and third components correspond to the payment and incurred data and the second component corresponds to the imposition of the restriction that ultimate claims for payments $P_{i,J}$ match incurred $I_{i,J}$ for all accident years, giving:

$$\begin{aligned} f\left(\mathbf{P}, \mathbf{I} \middle| \mathbf{\Theta}\right) \;=\; & \underbrace{\prod_{j=0}^{J}\prod_{i=0}^{J-j} \frac{1}{\sqrt{2\pi}\sigma_j P_{i,j}}\ \exp\left\{-\frac{1}{2\sigma_j^2}(\Phi_j - \log(\frac{P_{i,j}}{P_{i,j-1}}))^2\right\}}_{\text{Component1: payment}} \\ & \times \underbrace{\prod_{i=1}^{J} \frac{1}{\sqrt{2\pi(\upsilon_{J-i}^2 - \omega_{J-i}^2)}I_{i,J-i}}\ \exp\left\{-\frac{1}{2(\upsilon_{J-i}^2 - \omega_{J-i}^2)}(\mu_{J-i} - \eta_{J-i} - \log(\frac{P_{i,J-i}}{P_{i,J-i}}))^2\right\}}_{\text{Component2: Discounted final development year restricted payment and incurred}} \\ & \times \underbrace{\prod_{j=0}^{J-1}\prod_{i=0}^{J-j-1} \frac{1}{\sqrt{2\pi}\tau_j I_{i,j}}\ \exp\left\{-\frac{1}{2\tau_j^2}(-\Psi_j + \log(\frac{I_{i,j}}{I_{i,j+1}}))^2\right\}}_{\text{Component3: incurred}}. \end{aligned} \tag{4.2.2}$$

where $\upsilon_j^2 = \sum_{m=0}^{J}\sigma_m^2 + \sum_{n=j}^{J-1}\tau_n^2$; $\omega_j^2 = \sum_{m=0}^{j}\sigma_m^2$; $\eta_j = \sum_{m=0}^{j}\Phi_m$; and $\mu_j = \sum_{m=0}^{J}\Phi_m - \sum_{n=j}^{J-1}\varphi_n$.

As noted in Merz and Wuthrich. (2010), there are several consequences of the model assumptions made regarding the restriction $I_{i,J} = P_{i,J}$ which applies for all $i \in \{1, 2 \ldots, J\}$. The first is that this condition is sufficient to guarantee that the ultimate loss will coincide for both claims payments and incurred loss data. The second is that this model assumes that there is no tail development factor beyond the ultimate year J. However this could be incorporated into such models, see Merz and Wüthrich (2010).

Merz and Wuthrich. (2010) discuss the relationship between the proposed Independent Log-Normal PIC model and existing models in the literature for payment loss based reserving and incurred loss based reserving. In particular, Merz and Wuthrich. (2010) [Section 2.1 and 2.2] show that the resulting cumulative payments $P_{i,j}$, conditional on model parameters $\mathbf{\Theta}$, will satisfy the model proposed in Hertig (1985) and the incurred losses $I_{i,j}$, conditional on model parameters $\mathbf{\Theta}$, will satisfy the model proposed in Gogol (1993). Lemma 4.2.1 summarizes their results.

Lemma 4.2.1. *The relationships between consecutive payment development year losses in a given accident year is given conditionally according to*

$$\left[\log\left(\frac{P_{i,j}}{P_{i,j-1}}\right)\middle| P_{0:J,0:j-1},\mathbf{\Theta}\right]\sim\mathcal{N}\left(\Phi_j,\sigma_j^2\right),\ \forall j\geq 0 \tag{4.2.3}$$

in agreement with Hertig's model. With conditional moments given according to the Chain Ladder property as in Merz and Wuthrich. (2010, Lemma 5.2) by,

$$\mathbb{E}\left[P_{i,j}|P_{0:J,0:j-1},\mathbf{\Theta}\right]=P_{i,j-1}\exp\left(\Phi_j+\sigma_j^2/2\right). \tag{4.2.4}$$

Furthermore, conditional upon the model parameters $\mathbf{\Theta}$, for all $0\leq j<j+l\leq J$ the relationships between consecutive incurred losses in a given accident year are given in Merz and Wuthrich. (2010) [Proposition 2.2] according to

$$\left[\log\left(I_{i,j+l}\right)\right| I_{0:J,0:j-1},I_{i,J},\mathbf{\Theta}]\sim\mathcal{N}\left(\mu_{j+1}+\frac{\nu_{j+1}^2}{\nu_j^2}\left(\log(I_{i,j})-\mu_j\right),\nu_{j+1}^2(1-\nu_{j+1}^2/\nu_j^2)\right), \tag{4.2.5}$$

These results are consistent with the model assumptions of Gogol, and are derived using properties of multivariate normal distribution, see Lemma 2.1 in Merz and Wuthrich. (2010).

Furthermore, for all accident years $i\in\{1,2,\ldots,J\}$, the resulting conditional expected ultimate payment loss equals the expected ultimate incurred loss, given the model parameters $\mathbf{\Theta}$, and is expanded in terms of the model parameters according to Equation (4.2.6), which are given by Merz and Wuthrich. (2010, Equation 1.1) as,

$$E\left[P_{i,J}\right|\mathbf{\Theta}]=E\left[I_{i,J}\right|\mathbf{\Theta}]=\exp\left(\sum_{m=0}^{J}\Phi_m+\sigma_m^2/2\right). \tag{4.2.6}$$

4.3. Incorporating the Gaussian Copula into Paid-Incurred-Claims Models

This section discusses an important aspect of extending the original Log-Normal PIC model of Merz and Wuthrich. (2010). In particular, when this model was developed in the independent setting it was observed by the authors that the assumption of conditional independence between $\xi_{i,j}$ and $\zeta_{k,l}$ for all $i,j,k,l\in\{1,2,\ldots,J\}$ was not necessarily consistent with observations. In particular, they note that Quarg and Mack (2004) discovered evidence for strong linear correlation between incurred and paid ratios. In Section 3.1 we explore in detail a different approach to incorporate dependence structures into the Log-Normal PIC model. Some aspects of the new approach have subsequently been proposed in the literature, while others are novel developments proposed in our article. We note that in developing the extended models, the convenient properties of conjugacy in the Bayesian framework, which aids estimation, is often lost. Hence, after presenting the models we develop efficient state of the art statistical estimation strategies based on adaptive MCMC.

4.3.1. Dependence via Payment Loss Ratios and Incurred Loss Ratios (Model II). This section generalizes the model by Happ and Wuthrich (2011), which has a static covariance structure, see Happ and Wuthrich (2011, Figure 1.1). We use a Bayesian approach, based on results in Lemma 1.2 and Model Assumptions 4.3.1, to estimate the extended models. We use properties of the matrix-variate Wishart and Inverse Wishart distributions to develop a Gaussian copula based statistical model. The relevant matrix-variate distributional assumptions and properties are provided in Lemma 1.2 and Lemma 1.3.

Model Assumptions 4.3.1 (Dependent Payment-Incured Ratios: PIC Log-Normal (*Model II*))**.** *The model assumptions for the Gaussian copula PIC Log-Normal model involve:*

- *The random matrix $\Sigma_i \in \mathbb{R}^{(2J+1)\times(2J+1)}$ representing the covariance structure for the random vector constructed from log payment ratios $\left(\xi_{i,j} = \log\left(\frac{P_{i,j}}{P_{i,j-1}}\right)\right)$ and log incurred loss ratios $\left(\zeta_{i,j} = \log\left(\frac{I_{i,j}}{I_{i,j+1}}\right)\right)$ in the i-th development year, denoted by $\Xi_i = (\xi_{i,0}, \xi_{i,1}, \zeta_{i,1}, \xi_{i,2}, \zeta_{i,2}, \ldots, \xi_{i,J}, \zeta_{i,J})$, is assumed distributed according to an inverse Wishart distribution prior (see definition and properties in Lemma 1.2 and Lemma 1.3),*

$$\Sigma_i \sim \mathcal{IW}\left(\Lambda_i, k_i\right) \tag{4.3.1}$$

  where Λ_i is a $((2J+1)\times(2J+1))$ positive definite matrix and $k_i > 2J$.
- *Conditionally, given $\mathbf{\Theta} = (\Phi_0, \ldots, \Phi_J, \Psi_0, \ldots, \Psi_J)$ and the $(2J+1)\times(2J+1)$-dimensional covariance matrix Σ, we have:*
  - *The random matrix, constructed from log payment ratios $\left(\xi_{i,j} = \log\left(\frac{P_{i,j}}{P_{i,j-1}}\right)\right)$ and log incurred loss ratios $\left(\zeta_{i,j} = \log\left(\frac{I_{i,j}}{I_{i,j+1}}\right)\right)$, denoted by Ξ and comprised of columns $\Xi_i = (\xi_{i,0}, \xi_{i,1}, \zeta_{i,1}, \xi_{i,2}, \zeta_{i,2}, \ldots, \xi_{i,J}, \zeta_{i,J})$, is assumed distributed according to a matrix-variate Gaussian distribution $f_{\Xi}^{MVN}\left(\Xi | M, \Sigma, \Omega\right)$, see the definition and properties in Lemma 1.1. The sufficient matrices are then the $((2J+1)\times(J+1))$ mean matrix $M = [\mathbf{\Theta}', \ldots, \mathbf{\Theta}']$, column dependence given by $((2J+1)\times(2J+1))$ dimensional covariance matrix Σ and row dependence given by $((J+1)\times(J+1))$ dimensional matrix Ω. If $\Omega = \mathbb{I}_{J+1}$, the covariance of the vectorization of $\widetilde{\Xi} =$*

$Vec(\Xi)$ is

$$\widetilde{\Sigma} = \mathbb{C}ov\left(\widetilde{\Xi}\right) = \bigoplus_{i=0}^{J} \Sigma_i = \begin{pmatrix} \Sigma_0 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \Sigma_1 & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \Sigma_J \end{pmatrix}, \tag{4.3.2}$$

where it is assumed in the model in Happ and Wuthrich (2011) that $\Sigma_i = \mathbb{C}ov(\Xi_i) = \Sigma$. However, this need not be the case and it is possible to consider two extensions, the first in which $\mathbb{C}ov(\Xi_i)$ varied as a function of $i \in \{0, 1, \dots, J\}$ and the second being the most general of these model structures given by the assumption

$$\mathbb{C}ov\left(\widetilde{\Xi}\right) = \Sigma \otimes \Omega. \tag{4.3.3}$$

  - *For all accident years, $i \in \{0, 1, \dots, J\}$, the ultimate payment losses and incurred losses are equal a.s., $P_{i,J} = I_{i,J}$.*

- *The matrix $\widetilde{\Sigma}$ is positive definite and the components of $\mathbf{\Theta}$ are independent with prior distributions*

$$\Phi_i \sim \mathcal{N}\left(\phi_i, s_i^2\right) \text{ and } \Psi_j \sim \mathcal{N}\left(\psi_j, t_j^2\right), \tag{4.3.4}$$

  and hyper-prior distributions

$$s_i^2 \sim \mathcal{IG}\left(\alpha_i, \beta_i\right) \text{ and } t_j^2 \sim \mathcal{IG}\left(a_j, b_j\right), \tag{4.3.5}$$

  for all $i \in \{1, \dots, J\}$ and $j \in \{0, \dots, J\}$.

This model extends the model developed in Happ and Wuthrich (2011) which assumes that Σ is fixed and known with a tri-diagonal structure. The extension we consider generalizes the dependence structure to be unknown *a priori* and given an inverse Wishart prior for matrix $\widetilde{\Sigma}$, so it forms part of the inference given the data, in the Bayesian inference. In addition, unlike in Happ and Wuthrich (2011) where they assume $\Sigma = \Sigma_i, \forall i \in \{0, 1, \dots, J\}$, we also allow for variation in Σ_i across development years.

Given these model assumptions, we now consider two consequences of the proposed model structures for the dependence between the log payment ratios and the log incurred loss ratios given in Lemma 4.3.1 and Lemma 4.3.2.

Lemma 4.3.1. *Conditional upon Λ_i and k_i, for all i in $\{0, 1, \dots, J\}$, and given the marginal distributions for Σ_i follow $\Sigma_i \sim \mathcal{IW}\left(\Lambda_i, k_i\right)$ with Λ_i a $((2J+1) \times (2J+1))$ positive definite matrix and $k_i > 2J$, the joint distribution for the $((2J^2+3J+1) \times (2J^2+3J+1))$ covariance matrix*

$\widetilde{\Sigma}$ *for the vectorized matrix for* Ξ*, given by* $\widetilde{\Xi} = Vec(\Xi)$*, under the assumption of independendence between development years,*

$$\widetilde{\Sigma} = \mathbb{C}ov\left(\widetilde{\Xi}\right) = \bigoplus_{i=0}^{J} \Sigma_i = (\Sigma_0 \oplus \cdots \oplus \Sigma_J), \tag{4.3.6}$$

results in a joint distribution given by:

$$\widetilde{\Sigma} \sim \mathcal{IW}\left(\widetilde{\Lambda}, \widetilde{k}\right), \tag{4.3.7}$$

with degrees of freedom $\widetilde{k} = \sum_{i=0}^{J} k_i > 2J^2 + 3J$ *and scale matrix*

$$\widetilde{\Lambda} = \bigoplus_{i=0}^{J} \Lambda_i. \tag{4.3.8}$$

Furthermore, the joint prior mean and mode for the distribution of the random matrix $\widetilde{\Lambda}$ *are*

$$\begin{aligned} \mathbb{E}\left[\widetilde{\Sigma}|\widetilde{\Lambda}, \widetilde{k}\right] &= \frac{1}{\left(\sum_{i=0}^{J} k_i\right) - (2J^2 + 3J)} \widetilde{\Lambda}, \text{ and} \\ m\left(\widetilde{\Sigma}\right) &= \frac{1}{2J^2 + 3J + 1 + \sum_{i=0}^{J} k_i} \widetilde{\Lambda}. \end{aligned} \tag{4.3.9}$$

The proof of this result is a consequence of the results in Lemma 1.2, the model assumptions and the properties of an inverse Wishart distributions; see Gupta and Nagar (2000)[Chapter 3].

□

Remarks 4.3.2. *We can demonstrate that under the proposed model assumptions the selection of the factorized covariance structure in Lemma 4.3.1 produces Bayesian conjugacy in the joint posterior of the model parameters given observed payment losses and incurred losses.*

Remarks 4.3.3. *It is noted in Happ and Wuthrich (2011) and Lemma 4.3.1 that in formulating the likelihood structure for this dependent model it is more convenient to work with the one-to-one (invertible) transformation for the observed data defined marginally for the* i*-th development year according to*

$$[\boldsymbol{X}_i|\boldsymbol{\Theta}] = [B_i\Xi_i|\boldsymbol{\Theta}] \sim \mathcal{N}\left(B_iM_i, B_i\Sigma_iB_i'\right), \tag{4.3.10}$$

where M_i *is the* i*-th column of matrix* M *and* $\boldsymbol{X}_i \in \mathbb{R}^{2J+1}$ *defined by* $\boldsymbol{X}_i = [\log I_{i,0}, \log P_{i,0}, \log I_{i,1}, \log P_{i,1}, \ldots, \log I_{i,J-1}, \log P_{i,J-1}, \log I_{i,J}]$*. This results in the joint matrix variate Normal distribution for random matrix* $X = [\boldsymbol{X}'_0, \boldsymbol{X}'_1, \ldots, \boldsymbol{X}'_J]$ *of all observed losses for payment and incurred data given after vectorisation* $\widetilde{\boldsymbol{X}} = Vec\left(\boldsymbol{X}\right)$ *by*

$$\left[\widetilde{\boldsymbol{X}}|\boldsymbol{\theta}\right] = \left[B\widetilde{\Xi}|\boldsymbol{\Theta}\right] \sim \mathcal{N}\left(BVec(M), B\left(\Sigma \otimes \Omega\right)B^T\right). \tag{4.3.11}$$

Furthermore, if we consider the property of multivariate Gaussian distributions given in Lemma 4.3.2 we can find for the i-th accident year the required conditional distribution of the unobserved claims for payment and incurred loss data under the specified model. Furthermore, we can find the conditional distribution for unobserved claims for payment and incurred losses in the i-th accident year, given all observed claims triangles for payments and incurred losses data, see Lemma 4.3.2 below. This is directly relevant for specifying the resulting likelihood model.

Lemma 4.3.2. *Consider a $(n \times 1)$ random vector $\boldsymbol{Y}$ with multivariate Gaussian distribution, $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]$ and $\mathbb{C}ov(\boldsymbol{Y}) = \Sigma$, and partition $Y = \left[Y^{(1)\prime}, Y^{(2)\prime}\right]^{\prime}$. Then the conditional distribution of $Y^{(1)}$ given $Y^{(2)}$ and the marginal distribution of $\boldsymbol{Y}^{(1)}$ is*

$$\left[\boldsymbol{Y}^{(1)} | \boldsymbol{Y}^{(2)}\right] \sim \mathcal{N}\left(\bar{\boldsymbol{\mu}}, \bar{\Sigma}\right), \tag{4.3.12}$$

with $\bar{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}\left(\boldsymbol{Y}^{(2)} - \boldsymbol{\mu}^{(2)}\right)$ and the Schur complement $\bar{\Sigma} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}$ under the partitioning of the mean and covariance given by

$$\mu = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \text{ and } \Sigma = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{2,1} \\ \Sigma_{1,2} & \Sigma_{2,2} \end{pmatrix}. \tag{4.3.13}$$

Definition 4.3.4 below defines a family of permutation matrix operators. This permutation family allows the representation of the vectorization of the two loss triangles under different permutations that facilitate dependence specifications in the proposed models that admit conjugacy.

Definition 4.3.4. *Let Y be an $n \times n$ matrix, with $\widetilde{Y} = [Y_{1,1}, Y_{1,2}, \ldots]^{\prime}$ and with $Vec(Y)$ defined as $Vec(Y) = [Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n}, Y_{2,1}, \ldots, Y_{2,n}, \ldots, Y_{n,n}]^{\prime}$. Define the family of permutation matrix operators, denoted by $\mathcal{P}^*_{\boldsymbol{i}}$ and indexed by $p \times 2,\ p \leq n^2$, indices matrix (vector of tuple elements) $\boldsymbol{i}$ with j-th element $[\boldsymbol{i}\,]_j = \{(k,l)\,; k,l \in \{1,2,\ldots,n\}\}$, and defined according to the mapping $\mathcal{P}^*_{\boldsymbol{i}} : Vec(Y) \mapsto Vec(Y)^*$ given by*

$$\begin{aligned} \mathcal{P}^*_{\boldsymbol{i}}\ (Vec(Y)) &= P^*_{\boldsymbol{i}}\, Vec(Y) \\ &= \left[Y_{[\boldsymbol{i}\,]_1}, Y_{[\boldsymbol{i}\,]_2}, \ldots, Y_{[\boldsymbol{i}\,]_p}, Vec(Y)^{\prime}_{\backslash \boldsymbol{i}}\right]^{\prime}, \end{aligned} \tag{4.3.14}$$

where we define $Y_{[\boldsymbol{i}\,]_j}$ as the element of matrix Y corresponding to the resulting tuple index location in the j-th element (column) of (tuple vector) $\boldsymbol{i}$, P^*_i an $n^2 \times n^2$ permutation matrix defined by

$$P^*_{\boldsymbol{i}} = P_{\boldsymbol{i}} \oplus \mathbb{I}_{n^2-p} = \begin{bmatrix} P_{\boldsymbol{i}} & 0_{n^2-p,n^2-p} \\ 0_{n^2-p,n^2-p} & \mathbb{I}_{n^2-p} \end{bmatrix}, \tag{4.3.15}$$

and $P_{\boldsymbol{i}}$ is a matrix with only non-zero identity elements at the p locations in the indices matrix tuples in $\boldsymbol{i}$ corresponding to index elements.

Using the property of the multivariate Gaussian distribution in Lemma 4.3.2, one can state the result in Proposition 4.3.5 which is based on a generalization of the result in Happp and Wuthrich (2011)[Lemma 2.1] to the model developed above. We consider two cases for the dependence structures in Proposition 4.3.5 and Proposition 4.3.6.

Proposition 4.3.5. *Consider the i-th accident year. Conditional on the model parameters $\boldsymbol{\Theta}$ and the covariance matrix of the i-th accident year*

$$\Sigma_i = \begin{pmatrix} [\Sigma_i]_{1,1} & [\Sigma_i]_{2,1} \\ [\Sigma_i]_{1,2} & [\Sigma_i]_{2,2} \end{pmatrix}, \tag{4.3.16}$$

the dependence structure $\Omega = \mathbb{I}_{J+1}$ and the observed payment losses and incurred losses in the i-th accident year, denoted by $\boldsymbol{X}_i^{(1)} = [\log I_{i,0}, \log P_{i,0}, \log I_{i,1}, \log P_{i,1}, \ldots, \log I_{i,J-i}, \log P_{i,J-i}]$ with $\boldsymbol{X}_i \in \mathbb{R}^q$, the conditional distribution for the log of the unobserved payment losses and incurred losses
$(\boldsymbol{X}_i^{(2)} = [\log I_{i,J-i+1}, \log P_{i,J-i+1}, \ldots, \log I_{i,J-1}, \log P_{i,J-1}, \log I_{i,J}])$ is given by

$$\left[\boldsymbol{X}_i^{(2)} | \boldsymbol{X}_i^{(1)}, \boldsymbol{\Theta}\right] \sim \mathcal{N}\left(\bar{\boldsymbol{\mu}}^{(2)}, \bar{\Sigma}_i^{(2)}\right) \tag{4.3.17}$$

where $\bar{\boldsymbol{\mu}}_i^{(2)} = \boldsymbol{\mu}_i^{(2)} + [\Sigma_i]_{2,1} [\Sigma_i]_{1,1}^{-1} \left(\boldsymbol{X}_i^{(1)} - \boldsymbol{\mu}_i^{(1)}\right)$ and $\bar{\Sigma}_i^{(2)} = [\Sigma_i]_{22}$.

Proposition 4.3.6 (Conditional Distribution of Unobserved Payment and Incurred Losses)**.** *Consider the i-th accident year and define indices for this year (vector of tuples), given by matrix $\boldsymbol{i} = \{(k,j) : \forall j \in \{J-k+1, \ldots, J\}\} \cup \{(k,j) : \forall k \in \{0,1,\ldots,J\}, j \in \{0,\ldots,J-k\}\}$. Then consider the transformed vector of log payment and incurred losses $\mathcal{P}_{\boldsymbol{i}}^*\left(\widetilde{X}\right)$ defined by*

$$\mathcal{P}_{\boldsymbol{i}}^*\left(\widetilde{X}\right) \sim \mathcal{N}\left(P_{\boldsymbol{i}}^* Vec(M), P_{\boldsymbol{i}}^* \left(\Sigma \otimes \Omega\right) (P_{\boldsymbol{i}}^*)'\right), \tag{4.3.18}$$

for which the first $J-i$ elements of the permuted random vector $\left[\widetilde{X}^*\right]^{(1)} = \left[\mathcal{P}_{\boldsymbol{i}}^*\left(\widetilde{X}\right)\right]_{1:J-i-1}$ correspond to all un-observed payment and incurred loss random variables, and the remaining $J-i$ to $J-i+\left(\sum_{n=-1}^{J}(J-n)\right)$ elements are the observed payment and incurred data, denoted $\left[\widetilde{X}^*\right]^{(2)} = \left[\mathcal{P}_{\boldsymbol{i}}^*\left(\widetilde{X}\right)\right]_{J-i:J-i+\left(\sum_{n=-1}^{J}(J-n)\right)}$. Then, conditional on the model parameters $\boldsymbol{\Theta}$, the general dependence structre $\widetilde{\Sigma} = \Sigma \otimes \Omega$ with matrices Σ and Ω, and $\left[\widetilde{X}^*\right]^{(2)}$ the following results hold:

- *The conditional distribution for the log of the <u>unobserved payment losses and incurred losses</u> in the i-th year, corresponding to the first $J-i$ elements of the permuted random vector*

$\left[\widetilde{X}^{*}\right]^{(1)}=\left[\mathcal{P}^{*}_{\boldsymbol{i}}\left(\widetilde{X}\right)\right]_{1:J-i-1}$ *is given by*

$$\left[\left[\widetilde{X}^{*}\right]^{(1)} \mid \left[\widetilde{X}^{*}\right]^{(2)}, \boldsymbol{\Theta}\right] \sim \mathcal{N}\left(\bar{\boldsymbol{\mu}}^{(1)}, \bar{\Sigma}_i^{(1)}\right). \tag{4.3.19}$$

- *The covariance matrix* $\bar{\Sigma}_i^{(1)}$ *is the postive definite* $\left(J-i+\left(\sum_{n=-1}^{J}(J-n)\right)\right)\times\left(J-i+\left(\sum_{n=-1}^{J}(J-n)\right)\right)$ *sub-matrix denoted below by* Γ *and defined by the top sublock of the permuted covariance matrix*

$$P^{*}_{\boldsymbol{i}}\left(\Sigma\otimes\Omega\right)\left(P^{*}_{\boldsymbol{i}}\right)' = \begin{bmatrix} \Gamma & \left[P^{*}_{\boldsymbol{i}}\left(\Sigma\otimes\Omega\right)\left(P^{*}_{\boldsymbol{i}}\right)'\right]_{2,1} \\ \left[P^{*}_{\boldsymbol{i}}\left(\Sigma\otimes\Omega\right)\left(P^{*}_{\boldsymbol{i}}\right)'\right]_{1,2} & \left[P^{*}_{\boldsymbol{i}}\left(\Sigma\otimes\Omega\right)\left(P^{*}_{\boldsymbol{i}}\right)'\right]_{2,2} \end{bmatrix}. \tag{4.3.20}$$

- *Given, this covariance matrix one specifies the conditional mean vector, denoted by* $\bar{\boldsymbol{\mu}}^{(1)}=\boldsymbol{\mu}^{(1)}+\Gamma_{2,1}\Gamma_{1,1}^{-1}\left(\left[\widetilde{X}^{*}\right]^{(2)}-\boldsymbol{\mu}^{(2)}\right)$*, according to the subblocks of the* Γ *covariance matrix defined with respect to the first* $J-i$ *elements* $\left[\widetilde{X}^{*}\right]^{(1)}$ *and remaining elements of* $\left[\widetilde{X}^{*}\right]^{(2)}$ *as well as* $\boldsymbol{\mu}^{(1)}=\left[P^{*}_{\boldsymbol{i}}Vec(M)\right]_{1:J-i}$ *and the second* $J-i$ *to* $J-i+\left(\sum_{n=-1}^{J}(J-n)\right)$ *elements are given by* $\boldsymbol{\mu}^{(2)}=\left[P^{*}_{\boldsymbol{i}}Vec(M)\right]_{J-i:J-i+\left(\sum_{n=-1}^{J}(J-n)\right)}$.

Having specified these statistical assumptions, we can formulate the joint likelihood from the observed data for both payments and incurred claims conditional upon the model parameters according to Equation (4.3.21). The joint data likelihood function in the dependent Log-Normal PIC Model I for the random vector of observations corresponding to the first $\sum_{n=-1}^{J}(J-n)$ elements of the permuted random vector, given by $\left[\widetilde{X}^{*}\right]^{(1)}=\left[\mathcal{P}^{*}_{\boldsymbol{i}}\left(\widetilde{X}\right)\right]_{1:\left(\sum_{n=-1}^{J}(J-n)\right)}$, where we define indices in this case by $\boldsymbol{i}=\{(i,j):\forall i\in\{0,1,\ldots,J\}, j\in\{0,\ldots,J-i\}\}$. The resulting likelihood is given by the matrix-variate Gaussian distribution in Equation (4.3.21).

$$f\left(\left[\widetilde{X}^{*}\right]^{(1)}\middle|\boldsymbol{\Theta},\Sigma,\Omega\right)= \frac{\exp\left[\left(\left[\widetilde{X}^{*}\right]^{(1)}-\left[\mathcal{P}^{*}_{\boldsymbol{i}}\left(Vec(M)\right)\right]^{(1)}\right)\left[\left[P^{*}_{\boldsymbol{i}}\left(\Sigma\otimes\Omega\right)\left(P^{*}_{\boldsymbol{i}}\right)'\right]^{(1)}\right]^{-1}\left(\left[\widetilde{X}^{*}\right]^{(1)}-\left[\mathcal{P}^{*}_{\boldsymbol{i}}\left(Vec(M)\right)\right]^{(1)}\right)\right]}{(2\pi)^{\left(\sum_{n=-1}^{J}(J-n)\right)/2}\left|\left[P^{*}_{\boldsymbol{i}}\left(\Sigma\otimes\Omega\right)\left(P^{*}_{\boldsymbol{i}}\right)'\right]^{(1)}\right|^{\left(\sum_{n=-1}^{J}(J-n)\right)/2}} \tag{4.3.21}$$

We note that our proposed models also allow one to consider the dependence structures of Happ and Wuthrich (2011) who assume that $\Sigma_i=\Sigma, \forall i\in\{0,1,\ldots,J\}$ and $\Omega=\mathbb{I}_{J+1}$, with the specific setting of Σ via a tri-diagonal correlation matrix with three correlation parameters which are assumed either known *a priori* or estimated prior to inference in the PIC model. Such an approach was motivated by the belief that a positive change in incurred loss results in an immediate payment in the same development period, and the remaining increased expectation is paid with some settlement delay. Therefore, the incurred losses increments ζ_i^j are assumed to be positively correlated to the claims payments increments $\xi_{i,j}$, $\xi_{i,j+1}$ and $\xi_{i,j+2}$ with positive correlations ρ_0, ρ_1, ρ_2, respectively. However, the argument for more general dependence structure that were introduced as extensions to the model of Happ and Wuthrich (2011) are developed to account for the fact that these assumption may not be true, especially in long tail portfolios, such as compulsory third party. If the status of a claimant changes and requires long term medical treatment and rehabilitation, it might result in substantially high loss in the subsequent lengthy lag periods. The chapter also assumes that the dependence is the same across different lag years, which is not always a realistic feature of such data. Our article aims to fill this gap and enhance the correlation structure in PIC models whilst maintaining a parsimonious model specification.

4.3.2. Dependence Between Development Lag Years for Payment Losses and Incurred Losses (Model III). This section considers an alternative dependence structure motivated by the fact that dependence between lag years is practically appealing in claims reserving practice. It affects the estimation of outstanding claims the most, and is widely recognized by actuaries in claims reserving. Lag is the measure of the difference between incurred month and paid month. Depending on the nature of the portfolio, many insurance claims often have lengthy settlement periods due to various reasons such as late reported claims, judicial proceedings, or schedules of benefits for employer's liability claims. During the lengthy lag periods, large payments in the previous lag period normally follow by small payments in the subsequent lag period. There may in fact be positive correlation if all periods are equally impacted by a change in claims status, e.g. if a claim becomes litigated, resulting in a huge increase in claims cost. There may also be negative correlation if a large settlement in one period replaces a stream of payments in later periods. The model developed in this section mainly focuses on capturing this feature of dependence between lag years. To achieve this we propose a block covariance structure for the covariance matrix, which will respect the dependence between each lag point. The model we propose is summarised in Model Assumptions 4.3.7 below.

Model Assumptions 4.3.7 (Dependent Development Lag Years: PIC Log-Normal (*Model III*))**.** *The following statistical model assumptions are developed:*

- *Let* $\Sigma_i^P \in \mathbb{SD}^+(J-i)$ *be the* $(J-i) \times (J-i)$ *random covariance matrix on the space* $\mathbb{SD}^+(J-i)$ *of positive definite covariance matrices of dimension* $(J-i) \times (J-i)$ *corresponding to the observed payment data* $[\log P_{i,0}, \log P_{i,1}, \ldots, \log P_{i,J-i}]$ *in the* i*-th accident year and analogously for incurred loss data* $\Sigma_i^I \in \mathbb{SD}^+(J-i)$*. When* $i = 0$ *we consider* $\Sigma_0^P \in \mathbb{SD}^+(J+1)$ *and for incurred loss data* $\log I_{0,0:J-1}$ *with* $\Sigma_0^I \in \mathbb{SD}^+(J)$*.*

Assume an inverse Wishart distribution (see Lemma 1.3 and Lemma 1.2) for each matrix defined according to

$$\Sigma_i^P \sim \mathcal{IW}\left(\Lambda_i^P, k_i^P\right) \;\; \text{and} \;\; \Sigma_i^I \sim \mathcal{IW}\left(\Lambda_i^I, k_i^I\right), \tag{4.3.22}$$

where Λ_i^P and Λ_i^I are the inverse scale matrices for the prior for the payment and incurred loss data covariance priors respectively. Hence, the joint covariance between all observed payment and incurred loss data satisfies the telescoping diagonal block size covariance structure:

$$\begin{aligned}\widetilde{\Sigma} &= \mathbb{C}ov\left([\log P_{0,0}, \ldots, \log P_{0,J}, \log P_{1,0}, \log P_{1,J-1}, \ldots, \log P_{J,0}, \log I_{0,0}, \ldots, \log I_{0,J-1}, \ldots, \log I_{J,0}]\right) \\ &= \left(\bigoplus_{i=0}^{J} \Sigma_0^P\right) \oplus \left(\bigoplus_{i=0}^{J} \Sigma_0^I\right) \sim \mathcal{IW}\left(\left(\bigoplus_{i=0}^{J} \Lambda_0^P\right) \oplus \left(\bigoplus_{i=0}^{J} \Lambda_0^I\right), \sum_{i=0}^{J}\left(k_i^P + k_i^I\right)\right).\end{aligned} \tag{4.3.23}$$

- Conditionally, given $\mathbf{\Theta} = (\Phi_0, \ldots, \Phi_J, \Psi_0, \ldots, \Psi_J)$ and the covariance matrix $\widetilde{\Sigma}$, we have the following results
  - Consider the marginal distribution of the first $\left(\sum_{n=-1}^{J}(J-n)\right)$ elements of the vectorized random matrix of observed payment and incurred losses, with i-th column $\boldsymbol{X}_i \in \mathbb{R}^{2J+1}$ given by

    $$\boldsymbol{X}_i = [\log I_{i,0}, \log P_{i,0}, \log I_{i,1}, \log P_{i,1}, \ldots, \log I_{i,J-1}, \log P_{i,J-1}, \log I_{i,J}].$$

    Then given the matrix of permutation indices $\boldsymbol{i} = [(1,2),(1,4),\ldots,(1,2(J-1))$ $,(2,2),(2,4),\ldots,(2,2J-4),\ldots,(J,1),(1,1),(1,3),\ldots(J-1,1),(J-1,2)]$ characterizing the elements of the marginal distribution for the observations, the transform $\mathcal{P}^*_{\boldsymbol{i}}\,(Vec(\boldsymbol{X}))$ has multivariate Gaussian distribution with covariance structure $\widetilde{\Sigma}$. Note, $\mathcal{P}^*_{\boldsymbol{i}}\,(Vec(\boldsymbol{X})) = [\log P_{0,0}, \log P_{0,1}, \ldots, \log P_{0,J}$ $, \ldots, \log P_{J,0}, \log I_{0,0}, \ldots, \log I_{0,J-1}, \log I_{1,0}, \ldots, \log I_{J-1,0}, \log I_{J-1,1}]$.
  - For all accident years, $i \in \{0, 1, \ldots, J\}$, the ultimate payment losses and incurred losses are equal almost surely, $P_{i,J} = I_{i,J}$.
- The matrix $\widetilde{\Sigma}$ is positive definite and the components of $\mathbf{\Theta}$ are independent with prior distributions

$$\Phi_i \sim \mathcal{N}\left(\phi_i, s_i^2\right) \;\; \text{and} \;\; \Psi_j \sim \mathcal{N}\left(\psi_j, t_j^2\right) \tag{4.3.24}$$

and hyper-prior distributions

$$s_i^2 \sim \mathcal{IG}\left(\alpha_i, \beta_i\right) \;\; \text{and} \;\; t_j^2 \sim \mathcal{IG}\left(a_j, b_j\right) \tag{4.3.25}$$

for all $i \in \{1, \ldots, J\}$ and $j \in \{0, \ldots, J\}$.

This proposed model is therefore another generalization of the dependence structure of the model structure proposed in Happ and Wuthrich (2011). As such, the likelihood structure is given by the multivariate Gaussian given in Equation (4.3.21) with the covariance matrix given by the telescoping diagonal block size covariance matrix structure in Equation (4.3.23).

4.3.3. Hierarchical Bayesian Conjugacy Under Gausian Copula Dependent PIC: Models I, II, III. Under the Gaussian copula based dependence models, the ability to obtain the observed data likelihood in the form of a multivariate Gaussian distribution means that we obtain conjugacy properties. This makes the estimation of such models by MCMC more efficient because we can us Gibbs sampling in blocks. This section presents a generic set of such conjugate models for any of the dependence structures specified in Models I, II and III.

Lemma 4.3.3. *Conditional upon the parameters $\boldsymbol{\Theta}$ and the covariance matrix Σ, the permuted data $\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))$ can be transformed to produce the independent likelihood in Equation (4.2.2). This is achieved by considering the class of vector transformations $\mathcal{T} : \mathbb{R}^{(d\times 1)} \mapsto \mathbb{R}^{(d\times 1)}$, such that if the initial covariance structure of random vector $\boldsymbol{X}$ was given by $\Sigma = \mathbb{C}ov\left(\boldsymbol{X}\right)$, then the resulting covariance structure $\mathbb{C}ov\left(\mathcal{T}(\boldsymbol{X})\right) = \mathbb{I}_d$. The required rotation-dilation transformation is obtained by the spectral decomposition of the covariance according to a spectral decomposition (see Stoica and Moses (1997)) $\Sigma = U\Lambda^{\frac{1}{2}}U'$ where U is a $(d\times d)$ matrix of eigenvectors of Σ and Λ is a diagonal $d\times d$ matrix of the eigenvalues of Σ. Therefore the following holds for each of the models under a transform of the vector of permuted observations $\mathcal{T}\left(\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))\right)$:*

(1) *Model II - When $\tilde{\Sigma} = \Sigma\otimes\Omega$, with $\Omega = \mathbb{I}_{J+1}$ then, $\mathcal{T}\left(\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))\right) = \left(U\Lambda^{\frac{1}{2}}\otimes\mathbb{I}_{J+1}\right)\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X})$ where the $((2J+1)\times(2J+1))$ covariance Σ is decomposed as $U\Lambda^{\frac{1}{2}}U'$.*

(2) *Model II - When $\tilde{\Sigma} = \bigoplus_{i=0}^{J}\Sigma_i$, $\mathcal{T}\left(\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))\right) = \left(\bigoplus_{i=0}^{J}U_i\Lambda_i^{\frac{1}{2}}\right)\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))$, where each accident year's dependence between payments and incurred losses is given by the $(2J+1)\times(2J+1)$ matrix Σ_i which is decomposed as $U_i\Lambda_i^{\frac{1}{2}}U_i'$.*

(3) *Model III - When $\tilde{\Sigma} = \left(\bigoplus_{i=0}^{J}\Sigma_0^P\right)\oplus\left(\bigoplus_{i=0}^{J}\Sigma_0^I\right)$,*

$$\mathcal{T}\left(\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))\right) = \left(\bigoplus_{i=0}^{J}U_i^P\left(\Lambda_i^P\right)^{\frac{1}{2}}\right)\oplus\left(\bigoplus_{i=0}^{J}U_i^I\left(\Lambda_i^I\right)^{\frac{1}{2}}\right)\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))$$

where each of the covariance matrices Σ_i^P and Σ_i^I decomposed to $U_i^P\left(\Lambda_i^P\right)^{\frac{1}{2}}(U_i^P)'$ and $U_i^I\left(\Lambda_i^I\right)^{\frac{1}{2}}(U_i^I)'$.

In each case, the resulting transformed random vector $\mathcal{T}\left(\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))\right)$, with elements $\widetilde{P}_{i,j}$ and $\widetilde{I}_{i,j}$, will produce a likelihood model given for the transformed data according to the independent

Model I of Merz and Wuthrich. (2010) as defined in Equation (4.2.2). Of course this is defined now with respect to components in the likelihood corresponding to the transformed components, as detailed in Equation (4.3.11).

Remarks 4.3.8. *The consequence is that results in Lemma 4.3.3 are that the conjugacy properties derived for the independent model in Merz and Wuthrich. (2010) can be directly applied post-transformation. This is of direct interest for MCMC based sampling schemes.*

In the models described so far, the following full conditional posterior distributions are now of relevance to the Bayesian MCMC estimation procedures developed for Models I, II and III.

Lemma 4.3.4. *The full conditional posterior distributions for sub-blocks of the model parameters can be decomposed under Model I, II and III into a conjugate model.*

- ***Conjugate Posterior Distribution for Development Factors:*** *under the transformations $\mathcal{T}\left(\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))\right)$ on the data, described in Lemma 4.3.3, the full conditional posterior distributions for sub-blocks of the transformed model parameters $\left(\widetilde{\Phi}_{0:J}, \widetilde{\Psi}_{0:J}\right)$ are given by (see Merz and Wuthrich. (2010) [Theorem 3.4] for the independent case):*

$$\left[\widetilde{\Phi}_{0:J}, \widetilde{\Psi}_{0:J} | \Sigma, \Omega, \mathcal{T}\left(\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))\right)\right] \sim \mathcal{N}\left(\Pi, \Delta\right) \tag{4.3.26}$$

with posterior mean Π and posterior covariance Δ, where the components of $\Delta^{-1} = \left(a_{n,m}\right)_{0\geq n,m\leq 2J}$ are each given by

$$\begin{aligned} a_{n,m} &= \left(s_n^{-2} + (J-n+1)\sigma_n^{-2}\right)\delta_{n=m} + \sum_{i=0}^{(n-1)\wedge(m-1)} \left(\nu_i^2 - \omega_i^2\right)^{-1}, \ \textit{for } 0 \leq n,m \leq J, \\ a_{J+1+n,J+1+m} &= \left(t_n^{-2} + (J-n)\tau_n^{-2}\right)\delta_{n=m} + \sum_{i=0}^{n\wedge m} \left(\nu_i^2 - \omega_i^2\right)^{-1}, \ \textit{for } 0 \leq n,m \leq J-1, \\ a_{n,J+1+m} &= \Delta_{n,J+1+m} = -\sum_{i=0}^{(n-1)\wedge m} \left(\nu_i^2 - \omega_i^2\right)^{-1}, \ \textit{for } 0 \leq n \leq J, 0 \leq m \leq J-1; \end{aligned} \tag{4.3.27}$$

where $\delta_{n=m}$ is the indicator of the event that index m matches n, $m \wedge n$ is the minimum of m and n and the posterior mean is given on the transformed scale by,

$$\left[\widetilde{\Phi}_{0:J}, \widetilde{\Psi}_{0:J}\right] = \Delta\left(\widetilde{c}_0, \widetilde{c}_1, \ldots, \widetilde{c}_J, \widetilde{b}_0, \ldots, \widetilde{b}_J\right), \tag{4.3.28}$$

with

$$\begin{aligned}
\widetilde{c}_j &= s_j^{-2}\phi_j + \sigma_j^2 \sum_{i=0}^{J-j} \log\left(\frac{\widetilde{P}_{i,j}}{\widetilde{P}_{i,j-1}}\right) + \sum_{i=J-j+1}^{J} \left(\nu_{J-i}^2 - \omega_{J-i}^2\right)^{-1} \log\left(\frac{\widetilde{I}_{i,J-i}}{\widetilde{P}_{i,J-i}}\right), \\
\widetilde{b}_j &= t_j^{-2}\psi_j + \tau_j^2 \sum_{i=0}^{J-j-1} \log\left(\frac{\widetilde{I}_{i,j}}{\widetilde{I}_{i,j+1}}\right) - \sum_{i=J-j}^{J} \left(\nu_{J-i}^2 - \omega_{J-i}^2\right)^{-1} \log\left(\frac{\widetilde{I}_{i,J-i}}{\widetilde{P}_{i,J-i}}\right).
\end{aligned} \tag{4.3.29}$$

Given the transform vector $\left[\widetilde{\Phi}_{0:J}, \widetilde{\Psi}_{0:J}\right]$, the parameters on the orginal scale can be expressed according to the unique solution to the system of linear equations:

(1) *Model II - On the untransformed scale, the solution is given by the following system of equations*

$$[\Phi_{0:J}, \Psi_{0:J}]' = U^{-1}\Lambda^{-\frac{1}{2}} \left[\widetilde{\Phi}_{0:J}, \widetilde{\Psi}_{0:J}\right]. \tag{4.3.30}$$

(2) *Model II - On the untransformed scale, the solution is given by the following system of equations for each $i \in \{0, 1, \ldots, J\}$, where we can randomly select i or deterministically scan through i for the results,*

$$[\Phi_{0:J}, \Psi_{0:J}]' = U_i^{-1}\Lambda_i^{-\frac{1}{2}} \left[\widetilde{\Phi}_{0:J}, \widetilde{\Psi}_{0:J}\right]. \tag{4.3.31}$$

(3) *Model III - On the untransformed scale, the solution is given by the following system of equations,*

$$\begin{aligned}
[\Phi_{0:J}, \Phi_{0:J-1}, \Phi_{0:J-2}, \ldots, \Phi_J]' &= \bigoplus_{i=0}^{J} \left(U_i^P\right)^{-1} \left(\Lambda_i^P\right)^{-\frac{1}{2}} \left[\widetilde{\Phi}_{0:J}, \widetilde{\Phi}_{0:J-1}, \widetilde{\Phi}_{0:J-2}, \ldots, \widetilde{\Phi}_J\right], \\
[\Psi_{0:J}, \Psi_{0:J-1}, \Psi_{0:J-2}, \ldots, \Psi_J]' &= \bigoplus_{i=0}^{J} \left(U_i^I\right)^{-1} \left(\Lambda_i^I\right)^{-\frac{1}{2}} \left[\widetilde{\Psi}_{0:J}, \widetilde{\Psi}_{0:J-1}, \widetilde{\Psi}_{0:J-2}, \ldots, \widetilde{\Psi}_J\right].
\end{aligned}$$

- ***Conjugate Posterior Distribution for the Covariance Matrix:*** *Given the transformed observed payment and incurred losses have a multivariate Gaussian likelihood, as specified in Equation (4.3.21), with covaraince matrix $\widetilde{\Sigma} = \Sigma \otimes \Omega$ and mean vector $Vec(M)$. Then the posterior for the covariance matrix is the Inverse-Wishart-Gaussian distribution detailed in Peters et al. (2011b) [Section 3] and Peters et al. (2011c)*

$$\left[\widetilde{\Sigma}|\Phi_{0:J}, \Psi_{0:J}, \mathcal{T}\left(\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))\right)\right] \sim \mathcal{IW}\left(\Lambda + \mathcal{T}\left(\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))\right)\mathcal{T}\left(\mathcal{P}^*_{\boldsymbol{i}}\ (Vec(\boldsymbol{X}))\right)', dim\left(Vec(\boldsymbol{X})\right) +$$

In cases in which the covariance matrix $\widetilde{\Sigma}$ takes any of the block diagonal forms presented in Models II and III, we may utilise Lemma 1.2 and the result in Equation (4.3.4) to further decompose the posterior covariance into blockwise components.

- ***Conjugate Posterior Distribution for the Hyper-Parameters on Development Factors:*** *For all i we have the following Inverse Gamma-Gaussian conjugacy for the hyper parameters in Models II and III,*

$$\left[s_i^2|\Phi_i\right] \sim \mathcal{IG}\left(\alpha_i + \frac{1}{2}, \beta_i + \frac{\left(\Phi_i - \phi_i\right)^2}{2}\right) \quad \textit{and} \quad \left[t_i^2|\Psi_i\right] \sim \mathcal{IG}\left(a_i + \frac{1}{2}, b_i + \frac{\left(\Psi_i - \psi_i\right)^2}{2}\right).$$

We next present alternative tail dependence structures for the PIC model. Previous studies on claims reserving that have incorporated copula based models, such as Zhang et al. (2012) have done so through regression based frameworks. Zhang et al. (2012) develop a parametric copula model to account for dependence between various lines of insurance claims. Their paper considers a bivariate Gaussian copula model with marginal generalized linear models to capture the positive correlation between the two insurance lines. Our article significantly extends the dependence modelling capability of the PIC model structure remaining in the frameworks presented above. However, to do so requires the introduction of auxiliary variables to enable computation. The approach developed involves modifying the posterior distribution by embeding the target posterior distribution for the model parameters into a much higher dimensional support comprised of the original model parameters and the additional auxiliary variables. The reason for this expansion of the posterior dimensions will be come clear below and is in general known in Bayesian statistics as an auxiliary variable framework.

4.4. Incorporating Mixture-Archimedean Copula Dependence Structures into Paid-Incurred-Claims Models: Model IV

This section presents an alternative parameteric approach to modelling and capturing dependence and tail dependence in the PIC model structure which involves considering copula based models within the PIC reserving framework. The dependence can be considered over the following combinations such as:

(1) Independent accident years and dependence between payment losses over the development years;
(2) Independent accident years and dependence between incurred losses over the development years;
(3) Independent accident years and dependence jointly between payment and incurred losses over the development years via a mixture copula, hierarchical copula (HAC) as in Kurowicka and Joe (2010), or a vine copula (d-vine, canononical vine) e.g. Aas et al. (2009);
(4) Dependent accident years and independent development years for payment, incurred or both sets of losses.

Our article concentrates on the mixture copula model which allows for combinations of upper and lower tail dependence of different strengths. We detail the class of auxiliary variable methods known in statistics as Data Augmentation and demonstrate how this class of models can be combined into our modelling framework to allow for consistent use of copula models in the PIC framework. There are many variations that can be explored in this approach. We give one such approach for Model IV, Assumptions 4.4.2, that is directly comparable to that used for Model II in Assumption 4.3.1.

We present fundamental properties of members of the Archimedean family of copula that we consider when constructing mixture copula models in the PIC framework in the Appendix, see Lemma 2.1 for the characteristics of the Archimedean family of copulas and Lemma 2.2 for the required distribution and densities for three members of this family. In addition references Denuit et al. (2005),Aas et al. (2009),Embrechts (2009), Min and Czado (2010) and Patton (2009) provide more detail.

In Lemma 2.1 the property of associativity of Archimedean copula models is particularly useful in the PIC model framework as it allows us to obtain analytic expressions for the likelihood structure of the matrix-variate PIC model. This is particularly useful if one specifies the model as a hierarchical Archimedean Copula (HAC) construction.

We consider the following popular members of the Archimedean family of copula models, due to their analytic tractability, their non-zero tail dependence properties and their parsimonious parameterizations. In addition, generating random variates from these class of models is trivial given the generator for the member of the Archimedean family of interest. Lemma 2.2 in the appendix presents the three Archimedean copulas for Clayton, Gumbel and Frank copulas that we consider and their properties. We use the following notation for copula densities we consider on $[0,1]^d$, see Nelsen (2006b, Section 4.4.3, Table 4.4.1) and Lemma 2.2: the Clayton copula density is denoted by $c^C(u_1,...,u_n;\rho^C)$ with $\rho^C \in [0,\infty)$ the dependence parameter; the Gumbel copula density is denoted by $c^G(u_1,...,u_n;\rho^G)$ with $\rho^G \in [1,\infty)$ the dependence parameter; and the Frank copula density is denoted by $c^F(u_1,...,u_n;\rho^F)$ with $\rho^F \in \mathbb{R}/\{0\}$ the dependence parameter.

In addition, we also note that the properties of these copulas of interest include that the Clayton copula does not have upper tail dependence, however its lower tail dependence can be expressed as $\lambda_L = 2^{-1/\rho^C}$. The Gumbel copula does not have lower tail dependence, however its upper tail dependence of the Gumbel copula can be expressed as $\lambda_U = 2 - 2^{1/\rho^G}$. The Frank copula does not have upper or lower tail dependence.

In this class of copula dependence models we consider the marginal distribution of each log payment or log incurred loss as distributed according to a Gaussian distribution and the joint distribution vector is modelled via a mixture copula comprised of the above three components from the

Archimedean family. Such a copula construction will still produce a copula as shown in Lemma 4.4.1.

Lemma 4.4.1. *Consider copula distributional members $C_i(u_1, u_2, \dots, u_n) \in \mathcal{A}^n$, where $\mathcal{A}^n$ defines the space of all possible n-variate distributional members of the Archimedean family of copula models, specified in Lemma 2.2. Any finite mixture distribution constructed from such copula components that admit tractable density functions $c_i(u_1, u_2, \dots, u_n)$, denoted by $\tilde{c}(u_1, u_2, \dots, u_n) = \sum_{i=1}^{m} w_i c_i(u_1, u_2, \dots, u_n)$, such that $\sum_{i=1}^{m} w_i = 1$, is also the density of a copula distribution.*

The proof of Lemma 4.4.1 is provided in Appendix 3.

4.4.1. Understanding Bayesian Data Augmentation. The modeling framework of Data Augmentation in the Bayesian framework is typically invoked to deal with situations in which the likelihood evaluation is intractable to perform point-wise. This would make Bayesian inference in such a model also generally intractable. For example if one considers the generic likelihood $p(\boldsymbol{y}_{1:n}|\boldsymbol{\theta})$ with observation random vectors $\boldsymbol{Y}_{1:n}$, which can be evaluate point-wise as a function of parameter vector $\boldsymbol{\theta}$ with respect to a realization of the observation process $\boldsymbol{y}_{1:n}$.

In the setting we encounter in the PIC models, we can generically consider the data random vector observation is partitioned into two vector sub-components $\boldsymbol{Y} = \left[\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}\right]$, of which only one component, say $\boldsymbol{Y}^{(1)}$, is actually observed. Then evaluation of the likelihood pointwise for $\boldsymbol{\theta}$ given a realization of $\boldsymbol{Y}^{(1)}_{1:n}$ would require solving the integral in Equation 4.4.1

$$p\left(\boldsymbol{Y}^{(1)}_{1:n}|\boldsymbol{\theta}\right) = \int p\left(\boldsymbol{Y}^{(1)}_{1:n}|\boldsymbol{\theta}, \boldsymbol{Y}^{(2)}_{1:n}\right) p\left(\boldsymbol{Y}^{(2)}_{1:n}|\boldsymbol{\theta}\right) d\boldsymbol{Y}^{(2)}_{1:n}. \tag{4.4.1}$$

Generally, this integral will not admit a closed form solution. Therefore, the Bayesian Data Augmentation approach involves extending the target posterior $p\left(\boldsymbol{\theta}|\boldsymbol{Y}^{(1)}_{1:n}\right)$ which is intractable due to the intractability of the likelihood to a new posterior model on a higher dimensional space, in which the target distribution is a marginal as given in Equation 4.4.2

$$p\left(\boldsymbol{\theta}, \boldsymbol{Y}^{(2)*}_{1:n}|\boldsymbol{Y}^{(1)}_{1:n}\right) = \frac{p\left(\boldsymbol{Y}^{(1)}_{1:n}|\boldsymbol{\theta}, \boldsymbol{Y}^{(2)*}_{1:n}\right) p\left(\boldsymbol{Y}^{(2)*}_{1:n}|\boldsymbol{\theta}\right) p(\boldsymbol{\theta})}{p\left(\boldsymbol{Y}^{(1)}_{1:n}\right)} \tag{4.4.2}$$

where $\boldsymbol{Y}^{(2)*}_{1:n}$ are auxiliary random vectors with prior distribution $p\left(\boldsymbol{Y}^{(2)*}_{1:n}|\boldsymbol{\theta}\right)$, 'augmented' to the posterior parameter space to allow tractability of the posterior inference. This will be explained in detail for the PIC copula models below.

4.4.2. Data Augmentation in the Bayesian PIC Copula Models. Definition 4.4.1 gives some useful notation for the results that follow.

Definition 4.4.1 (Auxiliary Data for Data Augmentation)**.** *Consider the defined loss data under the one-to-one (invertible) transformation for the observed data given by the joint matrix for all observations and auxiliary variables given by* $X = [\boldsymbol{X}'_0, \boldsymbol{X}'_1, \ldots, \boldsymbol{X}'_J]$*. In this framework, the* i*-th accident year is defined according to,* $\boldsymbol{X}_i = [\log I_{i,0}, \log P_{i,0}, \log I_{i,1}, \log P_{i,1}, \ldots, \log I_{i,J-1}, \log P_{i,J-1}, \log I_{i,J}]$*. Consider the permutation of each vector of log payments and log incurred losses given by* $\widetilde{\boldsymbol{X}}_i = \mathcal{P}^*_{\boldsymbol{i}}\left(\boldsymbol{X}_i\right) = \left[\log P_{i,0}, \log P_{i,1}, \ldots, \log P_{i,J}, \log I_{i,0}, \log I_{i,1}, \ldots, \log I_{i,J-1}\right]$*. Now consider the further partition by the decomposition of observed log payment losses and unobserved log payment losses as well as these quantities for the incurred losses defined for the* i*-th accident year by,*

$$
\begin{aligned}
\widetilde{\boldsymbol{X}}_i &= \left[\widetilde{\boldsymbol{X}}^P_{i,obs}, \widetilde{\boldsymbol{X}}^P_{i,aux}, \widetilde{\boldsymbol{X}}^I_{i,obs}, \widetilde{\boldsymbol{X}}^I_{i,aux}\right] \\
&= \left[\widetilde{\boldsymbol{X}}^P_{0,i,obs}, \ldots, \widetilde{\boldsymbol{X}}^P_{J-i,i,obs}, \widetilde{\boldsymbol{X}}^P_{J-i+1,i,aux}, \ldots, \widetilde{\boldsymbol{X}}^P_{J,i,aux}, \widetilde{\boldsymbol{X}}^I_{0,i,obs}, \ldots, \widetilde{\boldsymbol{X}}^I_{J-i,i,obs}, \widetilde{\boldsymbol{X}}^I_{J-i+1,i,aux}, \ldots, \widetilde{\boldsymbol{X}}^I_{J-1,i,au} \right. \\
&= \left[\underbrace{\log P_{i,0}, \ldots, \log P_{i,J-i}}_{\textit{observed Payments}}, \underbrace{\log P_{i,J-i+1}, \ldots, \log P_{i,J}}_{\textit{unobserved Payments}}, \underbrace{\log I_{i,0}, \ldots, \log I_{i,J-i}}_{\textit{observed Incurred}}, \underbrace{\log I_{i,J-i+1}, \ldots, \log I_{i,J-1}}_{\textit{unobserved Incurred}}\right]'.
\end{aligned}
\tag{4.4.3}
$$

Therefore the total data matrix of losses is given by $\widetilde{X} = \left[\widetilde{\boldsymbol{X}}_0, \ldots, \widetilde{\boldsymbol{X}}_J\right]$*. Note, the introduction in this section of the notation subscripts* obs *and* aux *allows us to make explicit the fact that the upper triangle of log payment losses and the upper triangle of log incurred losses are un-observed quantities for these random variables, while the lower triangular regions for such losses are observed. We denote these random variables as auxiliary variables (augmented) to the observed data random variables to create a complete data set of all losses.*

By considering the unobserved data in the lower payment and incurred loss triangles as auxiliary variables to be jointly estimated along with the model parameters, we will demonstrate below that only under this approach is consistency ensured in the copula structure of the PIC model. However, we first make the following model assumptions about the statistical features of the PIC model.

The following assumptions illustrate a choice of copula models for the mixture from the Archimedean family. However, there are many related specifications and frameworks that can be explored in this context, be we leave that to future research.

Model Assumptions 4.4.2 (Data-Augmented Mixture Copula PIC (*Model IV*))**.** *The model assumptions and specifications for the copula model we develop involve:*

- *Let the random matrix $\Sigma_i \in \mathbb{R}^{(2J+1)\times(2J+1)}$ be the covariance for $\widetilde{\boldsymbol{X}}_i = \left[\widetilde{\boldsymbol{X}}^P_{i,obs}, \widetilde{\boldsymbol{X}}^P_{i,aux}, \widetilde{\boldsymbol{X}}^I_{i,obs}, \widetilde{\boldsymbol{X}}^I_{i,aux}\right]$ with $\widetilde{\boldsymbol{X}}_i \in \mathbb{R}^{2J+1}$ for all $i = 0, \ldots, J$. We assume that Σ is diagonal where*

$$\Sigma_{i,i} \sim \mathcal{IG}\left(\alpha_i, \beta_i\right), \ \forall i \in \{0, \ldots, J\}, \tag{4.4.4}$$

where α_i and β_i are the known hyper-parameters for shape and scale.

- ***MARGINAL DISTRIBUTION:** given $\boldsymbol{\Theta} = (\Phi_0, \ldots, \Phi_J, \Psi_0, \ldots, \Psi_J)$ and covariance matrices $\Sigma, \Omega \in \mathbb{R}^{(2J+1)\times(2J+1)}$ and $\boldsymbol{\rho}$, we assume the marginal distribution of the random matrix, of all log payments and log incurred losses $\widetilde{X}$, comprised of columns $\widetilde{\boldsymbol{X}}_i$ for the i-th accident year is matrix-variate Gaussian with density, defined as in Lemma 1.1, with the $(2J+1)\times(J+1)$ mean matrix $\widetilde{M} = [\boldsymbol{\Theta}', \ldots, \boldsymbol{\Theta}']$, column dependence given by $(2J+1)\times(2J+1)$ covariance matrix Σ and row dependence given by $(J+1)\times(J+1)$ matrix Ω. Here we only consider the case of $\Omega = \mathbb{I}_{J+1}$ for the marginal independent case.*

- ***DATA AUGMENTED PIC MIXTURE COPULA LIKELIHOOD:** Given $\widetilde{\boldsymbol{X}}^P_{0,aux}, \widetilde{\boldsymbol{X}}^P_{1,aux}, \ldots, \widetilde{\boldsymbol{X}}^P_{J-1,aux}$, $\widetilde{\boldsymbol{X}}^I_{0,aux}, \widetilde{\boldsymbol{X}}^I_{1,aux}, \ldots, \widetilde{\boldsymbol{X}}^I_{J-1,aux}$, $\boldsymbol{\Theta} = (\Phi_0, \ldots, \Phi_J, \Psi_0, \ldots, \Psi_J)$, covariance matrices $\Sigma, \Omega \in \mathbb{R}^{(2J+1)\times(2J+1)}$ and $\boldsymbol{\rho}$, the joint distribution of the random matrix ($\widetilde{X}$) of all log permuted payment and incurred losses is assumed (in this example) to be independent between accident years. For the i-th column (corresponding to i-th accident year), the joint distribution of all losses ($\widetilde{\boldsymbol{X}}_i$) is assumed to be hierarchical Archimedean Copula (HAC) mixture copula specified by distribution,*

$$\begin{aligned}\left[\widetilde{X}\right]_{\bullet,i} &\sim \tilde{C}_{\boldsymbol{\rho}_i}\left(F\left(\widetilde{\boldsymbol{X}}^P_{i,obs}, \widetilde{\boldsymbol{X}}^P_{i,aux}, \widetilde{\boldsymbol{X}}^I_{i,obs}, \widetilde{\boldsymbol{X}}^I_{i,aux}; [\boldsymbol{M}]_{\bullet,i}, \Sigma\right)\right)\\ &= \tilde{C}^P_{\boldsymbol{\rho}^P_i}\left(F\left(\widetilde{\boldsymbol{X}}^P_{i,obs}, \widetilde{\boldsymbol{X}}^P_{i,aux}; [\boldsymbol{M}]^P_{\bullet i}, \Sigma\right)\right)\tilde{C}^I_{\boldsymbol{\rho}^I_i}\left(F\left(\widetilde{\boldsymbol{X}}^I_{i,obs}, \widetilde{\boldsymbol{X}}^I_{i,aux}; [\boldsymbol{M}]^I_{\bullet i}, \Sigma\right)\right),\end{aligned} \tag{4.4.5}$$

with supper script P and I denote the components for the log payments and log incurred losses in the i-th development year respectively and the density is given by

$$\begin{aligned}&f\left(\widetilde{\boldsymbol{X}}^P_{i,obs}, \widetilde{\boldsymbol{X}}^P_{i,aux}, \widetilde{\boldsymbol{X}}^I_{i,obs}, \widetilde{\boldsymbol{X}}^I_{i,aux} | [\boldsymbol{M}]_{\bullet i}, \Sigma, \boldsymbol{\rho}^P_i, \boldsymbol{\rho}^I_i\right)\\ &= \tilde{c}^P_{\boldsymbol{\rho}^P_i}\left(F\left(\widetilde{\boldsymbol{X}}^P_{i,obs}, \widetilde{\boldsymbol{X}}^P_{i,aux}; [\boldsymbol{M}]^P_{\bullet i}, \Sigma\right)\right)\tilde{c}^I_{\boldsymbol{\rho}_i}\left(F\left(\widetilde{\boldsymbol{X}}^I_{i,obs}, \widetilde{\boldsymbol{X}}^I_{i,aux}; [\boldsymbol{M}]^I_{\bullet i}, \Sigma\right)\right)\prod_{j=1}^{2J+1}\phi(\widetilde{X}_{j,i}; M_{j,i}, \Sigma_{i,i}),\end{aligned} \tag{4.4.6}$$

where

$$\begin{aligned}\tilde{c}^S_{\boldsymbol{\rho}_i}\left(F\left(\widetilde{\boldsymbol{X}}^S_{i,obs}, \widetilde{\boldsymbol{X}}^S_{i,aux}; [\boldsymbol{M}]^S_{\bullet i}, \Sigma\right)\right) &= w_1 c^G_{\rho_i^{(G,S)}}\left(F_{1,i}\left(\widetilde{\boldsymbol{X}}^S_{1,i,obs}; \tilde{M}^S_{1,i}, \Sigma_{1,1}\right), \ldots, F_{J,i}\left(\widetilde{\boldsymbol{X}}^S_{J,i,aux}; \tilde{M}^S_{J,i}, \Sigma_{J,J}\right)\right)\\ &+ w_2 c^F_{\rho_i^{(F,S)}}\left(F_{1,i}\left(\widetilde{\boldsymbol{X}}^S_{1,i,obs}; \tilde{M}^S_{1,i}, \Sigma_{1,1}\right), \ldots, F_{J,i}\left(\widetilde{\boldsymbol{X}}^S_{J,i,aux}; \tilde{M}^S_{J,i}, \Sigma_{J,J}\right)\right)\\ &+ (1-w_1-w_2) c^C_{\rho_i^{(C,S)}}\left(F_{1,i}\left(\widetilde{\boldsymbol{X}}^S_{1,i,obs}; \tilde{M}^S_{1,i}, \Sigma_{1,1}\right), \ldots, F_{J,i}\left(\widetilde{\boldsymbol{X}}^S_{J,i,aux}; \tilde{M}^S_{J,i}, \Sigma_{J,J}\right)\right), \ S \in \{P, I\},\end{aligned}$$

and such that $w_1 + w_2 + (1 - w_1 - w_2) = 1$. This specifies a mixture of central, upper and lower tail dependence as denoted by the mixture of Archimedian copula models made up of Frank, Clayton and Gumbel members, such that for the source of data S, the copula parameters for each Archimedian family member is given by $\rho_i^{(G,S)} > 0$, $\rho_i^{(C,S)} > 1$ and $\rho_i^{(F,S)} \in \mathbb{R}/\{0\}$. Therefore the total conditional distribution corresponding to the likelihood model considered is given by,

$$f\left(\tilde{X}|M,\Sigma,\Omega,\boldsymbol{\rho}\right) = \underbrace{\prod_{i=0}^{J} \tilde{c}^P_{\boldsymbol{\rho}_i^P}\left(F\left(\widetilde{\boldsymbol{X}}^P_{i,obs},\widetilde{\boldsymbol{X}}^P_{i,aux};[\boldsymbol{M}]^P_{\bullet i},\Sigma\right)\right)\tilde{c}^I_{\boldsymbol{\rho}_i^I}\left(F\left(\widetilde{\boldsymbol{X}}^I_{i,obs},\widetilde{\boldsymbol{X}}^I_{i,aux};[\boldsymbol{M}]^I_{\bullet i},\Sigma\right)\right)}_{\text{Copula Dependence in Data Augmented PIC Likelihood}}$$
$$\times \underbrace{\frac{\exp\left(-\frac{1}{2}tr\left[\Omega^{-1}\left(\widetilde{X}-M\right)'\Sigma^{-1}\left(\widetilde{X}-M\right)\right]\right)}{(2\pi)^{(2J^2+3J+1)/2}\,|\Omega|^{(2J+1)/2}\,|\Sigma|^{(J+1)/2}}}_{\text{Marginal Distribution in Data Augmented Likelihood PIC Model}}. \tag{4.4.7}$$

- *Assume that the tail dependence features of the Data-Augmented copula PIC model are such that the dependence structure is homogeneous accross accident years, $\boldsymbol{\rho}^P = \boldsymbol{\rho}^P_i$ and $\boldsymbol{\rho}^I = \boldsymbol{\rho}^I_i$ for all $i \in \{0,1,2,\ldots,J\}$.*
- *Conditional on Σ, $\boldsymbol{\Phi} = [\Phi_0,\Phi_1,\ldots,\Phi_J]$ and $\boldsymbol{\Psi} = [\Psi_0,\Psi_1,\ldots,\Psi_J]$ the hierarchical prior distribution on the auxiliary payment data for the i-th accident year is given by a normal distribution, centered on the development year mean,*

$$\widetilde{\boldsymbol{X}}^P_{i,aux} \sim \mathcal{N}\left(\left[\Phi_{J-i+1},\Phi_{J-i+2},\ldots,\Phi_J\right],\Sigma^P_2\right). \tag{4.4.8}$$

The hierarchical prior distribution on the auxiliary incurred loss data for the i-th accident year is given by

$$\widetilde{\boldsymbol{X}}^I_{i,aux} \sim \mathcal{N}\left(\left[\Psi_{J-i+1},\Psi_{J-i+2},\ldots,\Psi_J\right],\Sigma^I_2\right), \tag{4.4.9}$$

with Σ_2 the lower portion of covariance Σ corresponding to the lower triangle matrix from $(J-i+1)$ through to J for all $i \in \{0,1,2,\ldots,J\}$.

- *For all accident years, $i \in \{0,1,\ldots,J\}$, the ultimate payment losses and incurred losses are equal a.s., $P_{i,J} = I_{i,J},\ \mathbb{P}-a.s.$*
- *The matrix $\widetilde{\widetilde{\Sigma}}$ is positive definite and components of $\boldsymbol{\Theta}$ are independent with prior distributions*

$$\Phi_i \sim \mathcal{N}\left(\phi_i, s_i^2\right) \text{ and } \Psi_j \sim \mathcal{N}\left(\psi_j, t_j^2\right) \tag{4.4.10}$$

and hyper-prior distributions

$$s_i^2 \sim \mathcal{IG}\left(\alpha_i, \beta_i\right) \;\; and \;\; t_j^2 \sim \mathcal{IG}\left(a_j, b_j\right) \tag{4.4.11}$$

for all $i \in \{1, \dots, J\}$ *and* $j \in \{0, \dots, J\}$.

- *The matrix* Σ *is distributed as* $\Sigma \sim \mathcal{IW}\left(\Lambda, k\right)$ *and the copula parameters are distributed as* $\rho^{G,P} \sim \mathcal{IG}\left(\alpha^G, \beta^G\right)$, $\rho^{C,P} \sim \mathcal{IG}\left(\alpha^C, \beta^C\right)$ *and* $\rho^{F,P} \sim \mathcal{N}\left(0, \sigma^F\right)$

Hence, we have made precise the auxilliary data scheme used in formulating the Data-Augmented-PIC model. In particular illustrating the importance of the role of the auxiliary data in evaluation of the model and estimation of the PIC claim development factors. Also we note we get indirectly via the data augmentation the distribution for the predicted payment and incurred Loss reserves.

Remarks 4.4.3. *The following remarks provide motivation for the Data-Augmentation and resulting incorporation of auxiliary payment and incurred Losses data.*

- *The use of data augmentation in the above model structure is critical in the PIC model formulation, since it allows one to ensure that the dependence structure considered (in this case a HAC-Mixture) is consistent both across accident years and across development years.*
  Note: *In the case of a linear dependence structure such as with a covariance / correlation matrix under a Gaussian Copula or Independent Copula model, such as those presented previously under Models I,II, III, we have that conditional distributions and marginal distributions are Gaussian. This means that the evaluation of the likelihood is analytic without the need for auxiliary variables.*
- *In order to evaluate the likelihood one has two choices, to evaluate the observed data likelihood (Equation (4.4.12)) or to evaluate the full data likelihood (Equation (4.4.7)).*
  - *The PIC copula model equivalent of Equation 4.4.2 is the observed data likelihood is given for the* i*-th accident year by*

$$\begin{aligned}
&p\left(\widetilde{\boldsymbol{X}}^P_{i,obs}, \widetilde{\boldsymbol{X}}^I_{i,obs} | \boldsymbol{\Theta}, \Sigma, \Omega, \boldsymbol{\rho}\right) \\
&\;= \int \cdots \int p\left(\widetilde{\boldsymbol{X}}^P_{i,obs}, \widetilde{\boldsymbol{X}}^I_{i,obs} | \boldsymbol{\Theta}, \Sigma, \Omega, \boldsymbol{\rho}, \widetilde{\boldsymbol{X}}^P_{i,aux}, \widetilde{\boldsymbol{X}}^I_{i,aux}\right) p\left(\widetilde{\boldsymbol{X}}^P_{i,aux}, \widetilde{\boldsymbol{X}}^I_{i,aux} | \boldsymbol{\Theta}, \Sigma, \Omega, \boldsymbol{\rho}\right) d\widetilde{\boldsymbol{X}}^P_{i,aux} d\widetilde{\boldsymbol{X}}^I_{i,aux} \\
&\;= \int \cdots \int \tilde{c}^P_{\boldsymbol{\rho}^P_i}\left(F\left(\widetilde{\boldsymbol{X}}^P_{i,obs}, \widetilde{\boldsymbol{X}}^P_{i,aux}; [\boldsymbol{M}]^P_{\bullet i}, \Sigma\right)\right) \tilde{c}^I_{\boldsymbol{\rho}^I_i}\left(F\left(\widetilde{\boldsymbol{X}}^I_{i,obs}, \widetilde{\boldsymbol{X}}^I_{i,aux}; [\boldsymbol{M}]^I_{\bullet i}, \Sigma\right)\right) \\
&\quad \times f^{MVN}_{\widetilde{X}_{i,aux}}\left(\widetilde{x}_{i,aux}; M_{i,aux}, \Sigma^P_2 \oplus \Sigma^I_2\right) f^{MVN}_{\widetilde{X}_i}\left(\widetilde{x}_i; M_i, \Sigma\right) \; d\widetilde{\boldsymbol{X}}^P_{i,aux} \; d\widetilde{\boldsymbol{X}}^I_{i,aux}
\end{aligned}$$

where matrix-variate Gaussian distributions $f_X^{MVN}()$ and f_X^{MVN} are as defined in Lemma 1.1
with $\widetilde{X}_{i,aux} = Vec\left(\widetilde{\boldsymbol{X}}^{P}_{i,aux}, \widetilde{\boldsymbol{X}}^{I}_{i,aux}\right)$, $M_{i,aux} = Vec\left([\Phi_{J-i+1:J}]', [\Psi_{J-i+1:J-1}]'\right)$, $\widetilde{X}_i = [\widetilde{\boldsymbol{X}}^{P}_{i,obs}, \widetilde{\boldsymbol{X}}^{P}_{i,aux}, \widetilde{\boldsymbol{X}}^{I}_{i,obs}, \widetilde{\boldsymbol{X}}^{I}_{i,aux}]$ *and* $M_i = [\Phi_0, \ldots, \Phi_J, \Psi_0, \ldots, \Psi_J]$ *the equivalent mean.*

  - *Clearly, the marginalization required to evaluate the Observed data likelihood involves intractable integration, except in special cases in which the copula models are Gaussian or independence copulas.*

- *The full data likelihood comprised of observed and auxiliary data involves incorporating auxiliary variables to represent the unobserved data in the lower reserve triangle for payment and incurred loss triangles. These become part of the inference procedure and are required to be estimated jointly with the model parameters in the estimation methodology.*

4.5. Estimation via Adaptive Data-Augmented MCMC for Claims Reserving PIC Models

It has been shown for the Independent and Gaussian copula models that we can obtain the observed data likelihood analytically. Therefore the posterior distribution for all the model parameters can be sampled via a MCMC procedure comprised of block Gibbs sampler updates. In the case of a more general copula dependence model in which the observed data likelihood cannot be analytically evaluated pointwise, we must resort to a Data Augmentation scheme. In this case we will be able to perform sampling via a general MCMC Metropolis-Hastings sampler. In particular we will consider automating such a sampler using an adaptive MCMC scheme.

4.5.0.1. *Adaptive Metropolis within Data-Augmented Copula PIC Models.* This section presents the adaptive proposal we use to sample the parameters and the auxiliary variables. The advantage of an adaptive MCMC mechanism is that it automates the proposal design through consideration of a proposal distribution that learns the regions in which the posterior distribution for the static model parameters and auxiliary data has most mass. As such, the probability of acceptance under such an on-line adaptive proposal is likely to improve as the iterations progress and the generated MCMC samples will ideally have reduced autocorrelation. In such cases the variance of Monte Carlo estimators of integrals of smooth functionals formed from such samples will be reduced.

There are several classes of adaptive MCMC algorithms, see Roberts and Rosenthal. (2009). The distinguishing feature of adaptive MCMC algorithms, compared to standard MCMC, is the generation of the Markov chain via a sequence of transition kernels. Adaptive algorithms utilize a combination of time or state inhomogeneous proposal kernels. Each proposal in the sequence is allowed to depend on the past history of the Markov chain generated, resulting in many possible variants.

Haario et al. (2005b) develop an adaptive Metropolis algorithm with proposal covariance adapted to the history of the Markov chain was developed. Andrieu and Thoms. (2008) is presenting a tutorial discussion of the proof of ergodicity of adaptive MCMC under simpler conditions known as *Diminishing Adaptation* and *Bounded Convergence*. We note that when using inhomogeneous Markov kernels it is particularly important to ensure that the generated Markov chain is ergodic, with the appropriate stationary distribution. Two conditions ensuring ergodicity of adaptive MCMC are known as *Diminishing Adaptation* and *Bounded Convergence*. These two conditions are summarised by the following two results for generic Adaptive MCMC strategies on a parameter vector $\boldsymbol{\theta}$. As in Roberts and Rosenthal. (2009), we assume that each fixed MCMC kernel Q_γ, in the sequence of adaptions, has stationary distribution $P(\cdot)$ which corresponds to the marginal posterior of the static parameters. Define the convergence time for kernel Q_γ when starting from a state $\boldsymbol{\theta} \in E$, as $M_\epsilon(\boldsymbol{\theta}, \gamma) = \inf\{s \geq 1 : \|Q_\gamma^s(\boldsymbol{\theta}; \cdot) - P(\cdot)\| \leq \epsilon$. Under these assumptions, they give the following two conditions which are sufficient to guarantee that the sampler produces draws from the posterior distribution as the number of iterates tend to infinity. The two sufficient conditions are:

- *Diminishing Adaptation:* $\lim_{n\to\infty}\sup_{\boldsymbol{\theta}\in E}\|Q_{\Gamma_{s+1}}(\boldsymbol{\theta}, \cdot) - Q_{\Gamma_s}(\boldsymbol{\theta}, \cdot)\|_{tv} = 0$ in probability. Note, Γ_s are random indices.
- *Bounded Convergence:* For $\epsilon > 0$, the sequence $\{M_\epsilon(\boldsymbol{\theta}, \Gamma_j)\}_{j=0}^{\infty}$ is bounded in probability.

The sampler converges asymptotically in two senses,

- *Asymptotic convergence:* $\lim_{j\to\infty}\|\mathcal{L}\text{aw}(\boldsymbol{\theta}) - P(\boldsymbol{\theta})\|_{tv} = 0$ in probability.
- *Weak Law of Large Numbers*: $\lim_{j\to\infty}\frac{1}{j}\sum_{i=1}^{j}\phi(\boldsymbol{\theta}) = \int \phi(\boldsymbol{\theta})P(d\boldsymbol{\theta})$ for all bounded $\phi : E \to R$.

In general, it is non-trivial to develop adaption schemes which can be verified to satisfy these two conditions. In this chapter we use the adaptive MCMC algorithm to learn the proposal distribution for the static parameters in our posterior $\boldsymbol{\Phi}$. In particular we work with an adaptive Metropolis algorithm utilizing a mixture proposal kernel known to satisfy these two ergodicity conditions for unbounded state spaces and general classes of target posterior distribution, see Roberts and Rosenthal. (2009) for details.

4.5.0.2. *Euclidean and Riemann-Manifold Adaptive Metropolis within Data-Augmented Copula PIC Models.* This section presents the specific details of the Adaptive Metropolis algorithm that we combine with Data-Augmentation to obtain an MCMC sampler for the Data Augmented

Mixture Copula PIC Model proposed. This involves specifying the details of the proposal distribution in the AdMCMC algorithm which samples a new proposed update vector $\boldsymbol{\Upsilon}^*$ and matrix $\widetilde{\Sigma}^*$ from an existing Markov chain state $\boldsymbol{\Upsilon}$ with

$$\boldsymbol{\Upsilon} = \left[\boldsymbol{\Phi}, \boldsymbol{\Psi}, s^2_{0:J}, t^2_{0:J}, \boldsymbol{\rho}, \widetilde{\boldsymbol{X}}^P_{1,aux}, \ldots, \widetilde{\boldsymbol{X}}^P_{J,aux}, \widetilde{\boldsymbol{X}}^I_{1,aux} \ldots, \widetilde{\boldsymbol{X}}^I_{J,aux}\right]$$

and matrix $\widetilde{\Sigma}$. At the j-th iteration of the Markov chain we have existing state $\boldsymbol{\Upsilon}^{(j-1)}$ and $\widetilde{\Sigma}^{(j-1)}$ which is used to construct the proposal distribution $q\left(\boldsymbol{\Upsilon}^{(j-1)}, \boldsymbol{\Upsilon}^*\right) q\left(\widetilde{\Sigma}^{(j-1)}, \widetilde{\Sigma}^*\right)$. The choices we make for the two proposals will involve a novel development of a new adaptive proposal for positive definite matrices, required for the covariance matrix $\widetilde{\Sigma}$ should we choose not to specify it as diagonal.

Euclidean Space Adaptive Metropolis for Static Parameters:

We first detail the proposal for updating $\boldsymbol{\Upsilon}$ using a mixture of multivariate Gaussian distributions as specified for an Adaptive Metropolis algorithm which involves sampling from the proposal

$$q\left(\boldsymbol{\Upsilon}^{(t-1)}, \cdot\right) = w_1 \mathcal{N}\left(\boldsymbol{\Upsilon}; \boldsymbol{\Upsilon}^{(t-1)}, \frac{(2.38)^2}{d}\mathbb{C}\text{ov}\left(\left\{\boldsymbol{\Upsilon}^{(j)}\right\}_{0\leq j\leq t-1}\right)\right) + (1-w_1)\mathcal{N}\left(\boldsymbol{\Upsilon}; \boldsymbol{\Upsilon}^{(t-1)}, \frac{(0.1)^2}{d} I_{d,d}\right), \tag{4.5.1}$$

where we define the sample covariance for Markov chain past history by $\mathbb{C}\text{ov}\left(\left\{\boldsymbol{\Upsilon}^{(j)}\right\}_{0\leq j\leq t-1}\right)$ and we note the following recursive evaluation, which significantly aids in algorithmic computational cost reduction

$$\begin{aligned}
\mathbb{E}\left(\left\{\boldsymbol{\Upsilon}^{(j)}\right\}_{0\leq j\leq t}\right) &= \mathbb{E}\left(\left\{\boldsymbol{\Upsilon}^{(j)}\right\}_{0\leq j\leq t-2}\right) + \frac{1}{t}\left(\boldsymbol{\Upsilon}^{(t-1)} - \mathbb{E}\left(\left\{\boldsymbol{\Upsilon}^{(j)}\right\}_{0\leq j\leq t-1}\right)\right)\\
\mathbb{C}\text{ov}\left(\left\{\boldsymbol{\Upsilon}^{(j)}\right\}_{0\leq j\leq t}\right) &= \frac{1}{t+1}\left(\left(\boldsymbol{\Upsilon}^{(t-1)} - \mathbb{E}\left(\left\{\boldsymbol{\Upsilon}^{(j)}\right\}_{0\leq j\leq t}\right)\right)\left(\boldsymbol{\Upsilon}^{(t-1)} - \mathbb{E}\left(\left\{\boldsymbol{\Upsilon}^{(j)}\right\}_{0\leq j\leq t}\right)\right)' - \mathbb{C}\text{ov}\left(\left\{\boldsymbol{\Upsilon}^{(j)}\right\}_{0\leq}\right.\right.\\
&\quad \left. + \mathbb{C}\text{ov}\left(\left\{\boldsymbol{\Upsilon}^{(j)}\right\}_{0\leq j\leq t-1}\right)\right.
\end{aligned} \tag{4.5.2}$$

The theoretical motivation for the recommended choices of scale factors 2.38, 0.1 and dimension d are provided in Rosenthal et al. (2008).

Riemannian Manifold Adaptive Metropolis for Covariance Matrices:

Next we develop a novel proposal distribution for the sampling of the covariance matrix $\widetilde{\Sigma} \in \text{Sym}^+(d)$ in an adaptive MCMC proposal, restricted to the Riemann manifold of symmetric, postive definite $(d \times d)$ matrices, denoted by the space $\text{Sym}^+(d)$.

Remarks 4.5.1. *First, we note two properties of the marginal posterior* $p\left(\widetilde{\Sigma}\middle| \left\{\widetilde{X}^P_{i,obs}, \widetilde{X}^I_{i,obs}\right\}_{0\leq i\leq J}\right)$*: its distribution is restricted to the Riemann-manifold of symmetric positive definite matrices, but in general will not be Inverse-Wishart; second, the Markov chain samples drawn from this marginal distribution at iteration* t, $\left\{\widetilde{\Sigma}^{(s)}\right\}_{0\leq s\leq t}$*, are not independent. The consequence of this is that we* *cannot simply apply the property of closure under convolution of independent Wishart distributed random matrices to find a suitable proposal.*

Therefore, we will adopt a strategy to perform adaptive moment matching of a distribution with support $\mathrm{Sym}^+(d)$. We detail one possibility involving an inverse Wishart distribution fitted to the sample mean of the marginal posterior for the covariance. We note that future work could also consider specifying a distribution on the superset of the Riemannian manifold of symmetric positive definite matrices, given by the Riemannian manifold of symmetric matrices $\mathrm{Sym}^+(d) \subset \mathrm{Sym}(d)$.

Adaptive Metropolis inverse Wishart Mixture: We note that one way to achieve this is a mixture of inverse Wishart distributions given by

$$q\left(\widetilde{\Sigma}^{(t-1)}, \cdot\right) = w_1 \mathcal{IW}\left(\widetilde{\Sigma}; \Lambda_t^{adap}\left(\left\{\widetilde{\Sigma}^{(s)}\right\}_{0 \le s \le t-1}\right), p\right) + (1-w_1)\,\mathcal{IW}\left(\widetilde{\Sigma}; \Lambda, p\right). \quad (4.5.3)$$

Here, the adaptive proposal mixture component is specified through fixing the degrees of freedom p and then selecting Λ_t^{adap} with respect to the sample average of the covariance matrices $\left\{\widetilde{\Sigma}^{(s)}\right\}_{0 \le s \le t-1}$ which are samples from the matrix-variate marginal posterior in the Markov chain, thereby adapting the proposal to the Markov chain history. To perform the moment matching (Equation (4.5.4)), we note that we need to ensure that the sample average considered is restricted to the Riemann-manifold of positive definite matrices.

$$\Lambda_t^{adap}\left(\left\{\widetilde{\Sigma}^{(s)}\right\}_{0 \le s \le t-1}\right) = \widehat{\widetilde{\Sigma}}^{(t-1)}\left(p - dim(\widetilde{\Sigma}) - 1\right). \quad (4.5.4)$$

This is satisfied through the choice of the estimator

$$\widehat{\widetilde{\Sigma}}^{(t-1)} = \frac{1}{t-1}\sum_{s=1}^{t-1} \widetilde{\Sigma}^{(s)}. \quad (4.5.5)$$

To see this we observe that since we only form positive linear combinations of matrices on this manifold, with a scaling, such linear combinations will always remain on the manifold $\mathrm{Sym}^+(d)$.

4.6. Real Data Analysis

To illustrate the proposed models and compare with existing models and estimation methods in the actuarial literature we consider, as in Merz and Wuthrich. (2010), the example presented in Dahms (2008) and Dahms et al. (2009) (Tables 10 and 11). As in the second analysis framework in Merz and Wuthrich. (2010), we treat the claim development factors, the likelihood dependence parameters and the hyperparameters on the claim development factor priors as parameters which we incorporate into the posterior inference.

We present two sets of results, the first studies the performance of the adaptive Markov chain Monte Carlo algorithms developed for the estimation and inference of the posterior distributions for the PIC-Copula models for Gaussian Copula (Models III) and the Data-Augmented-Mixture-Copula PIC (Models IV). The second stage of results assesses the estimation of predictive distributions and dependence features of the PIC claims reserving models compared to the independent PIC Model, the payment only model and the incurred only models. In particular, we focuss analysis on the data sets studied in Merz and Wuthrich. (2010) for comparison of the influence of dependence features in PIC models versus independence assumptions when performing PIC claims reserving.

Convergence Analysis: In all the Markov chain Monte Carlo simulations, for each model (payment, payment-incurred Gaussian copula Model III; and Data-Augmented hierarchical Archemdean mixture copula Model IV), we carried out convergence diagnostics. This included the Gelman-Rubin R-statistics (all less than 1.5), the ACF plots for each parameter were checked to ensure all parameters had ACF's which were less than 10% by lag 20. Then the first 20% of samples were discarded as burnin and the remaining samples were used in inference results presented below.

4.6.1. Results: Euclidean and Riemann-Manifold Adaptive Metropolis for hierarchical Bayesian Copula PIC Models.

In the simualtion results, we consider a block Gibbs sampler with the following three stages performed at each iteration of the adaptive Metropolis-within-Gibbs sampler for the PIC Model III and Model IV:

> **Stage 1:** Perform exact sampling of the development factors and their hyperparameters under the conjugacy results developed.
>
> **Stage 2:** Perform Euclidean space Adaptive Metropolis updates of the Augmented Data variables using proposal in Equation (4.5.1).
>
> **Stage 3:** (Gaussian Copula Model III) - Perform Riemannian space Adaptive Metropolis updates of the covariance matrix in the Gaussian copula. Note, we consider the constrained specifications presented in the "Dependent Lag Years" model specification in Section 4.3.2, Equation (4.3.23). Under this hierarchical Bayesian model, the joint covariance between all observed payment and incurred loss data under the dependent development years assumption, satisfies a telescoping diagonal block size form covariance matrix structure. Hence, the sampling of this structure can be performed blockwise on each covariance sub-block;
> (Mixture Clayton-Gumbel Copula Model IV) - Perform Euclidean space Adaptive Metropolis updates of the mixture copula parameters.

4.6.1.1. *Hierarchical Bayesian Gaussian Copula (telescoping block covariance) PIC (Model III).* This section presents the estimation results for the Gaussian Copula based PIC models (Model III) on the real data. Figure 4.2 summarizes the dependence structure by a heatmap for the posterior distribution of the Gaussian copula covariance matrix. As mentioned in the introduction, the telescoping block covariance refers to the fact that the covariance structure is reducing in rank by 1 on each diagonal block for the payment data and then the incurred data. This model has the joint covariance between all observed payment and incurred loss data under the assumption that the development years are dependent, satisfying a telescoping diagonal block size form covariance matrix structure. Summarising the information from such posterior samples for distributions of covariance matrices is non-trivial as discussed in Tokuda et al. (2011), where they develop a four layer approach. Our article adopts aspects of the ideas proposed in Tokuda et al. (2011) to interpret the features of the posterior distribution samples for the dependence structures.

The posterior mean for estimated PIC covariance structure is obtained by using Monte Carlo samples from the Riemann-Manifold Adaptive Metropolis sampler and given by the estimator,

$$\mathbb{E}\left[\tilde{\Sigma}|\boldsymbol{P},\boldsymbol{I}\right] = \frac{1}{S}\sum_{s=1}^{S}\left\{\left(\bigoplus_{i=0}^{J}\Sigma_i^P\right)\oplus\left(\bigoplus_{i=0}^{J}\Sigma_i^I\right)\right\}^{(s)}, \tag{4.6.1}$$

where $\left\{\left(\bigoplus_{i=0}^{J}\Sigma_0^P\right)\oplus\left(\bigoplus_{i=0}^{J}\Sigma_0^I\right)\right\}^{(s)}$ is the s-th sample of the $J(J-1)\times J(J-1)$ covariance matrix. The estimated posterior mean covariance matrix is reported in a heatmap for the correlation matrix in Figure 4.2. In Figure 4.2, **Top panel:** Heatmap of the posterior distribution for the Gaussian copula covariance matrix (100×100), summarised by the heat map for the mean of correlation structure using samples from the Riemannian Manifold Adaptive Metropolis sampler under restriction to a telescoping diagonal block form. **Bottom Left Panel:** Heatmap for the posterior distribution sub-block covariance matrices Σ_0^P and Σ_1^P converted to correlation matrices. **Bottom Right Panel:** Heatmap for the posterior distribution sub-block covariance matrices Σ_0^I and Σ_1^I converted to correlation matrices. The color key is given at the top left. In addition, we present examples based on posterior mean covariance for covariance sub-blocks $p\left(\Sigma_4^P|\boldsymbol{P},\boldsymbol{I}\right)$ and then for $p\left(\Sigma_4^I|\boldsymbol{P},\boldsymbol{I}\right)$, where $\Sigma_4^P\in SP^+(6)$ and $\Sigma_4^I\in SP^+(5)$, again converted to heatmaps of the correlation. We see that although the priors selected for the dependence features in Model III in all cases favoured independence, since the scale matrices were all diagonal i.e. $\Lambda_5^P=\mathbb{I}_6$ and $\Lambda_4^I=\mathbb{I}_5$, the resulting summaries of the marginal posteriors of the covariances clearly indicate non-trivial dependence patterns in the development years within the payments data and the incurred loss data. This is observed throughout each sub-block covariance matrix.

Table 4.9 provides a second summary of the posterior for the covariance matrix which further demonstrates features of the dependence properties in the payment and incurred data per accident year and involves the estimates of the largest eigenvalue of each block diagonal matrix for the

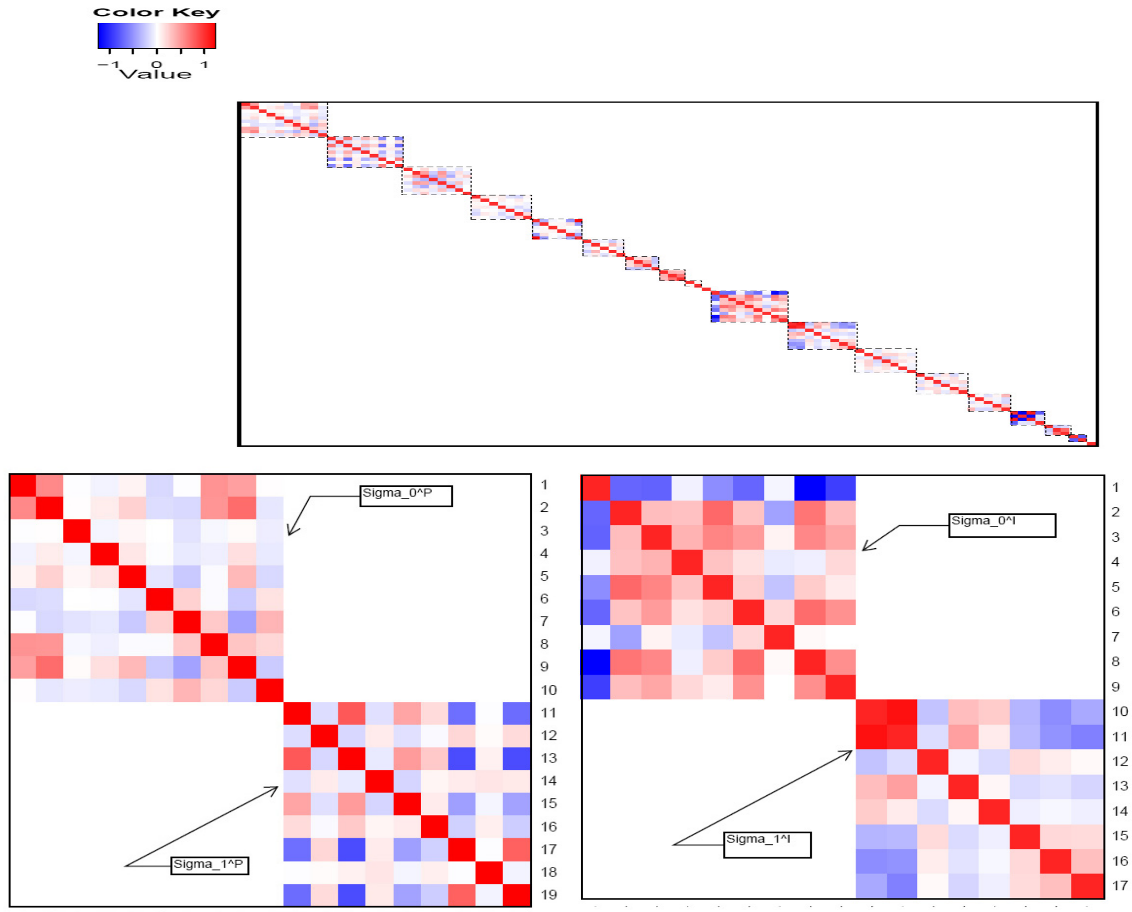


FIGURE 4.2. Heatmap of the posterior distribution

payment and incurred data as summary statistics. These estimates are given by

$$\widehat{\lambda}_i^{(s)} = \arg\max\left(\det(\Sigma_i^{(s)} - \lambda\mathbb{I}) = 0\right). \tag{4.6.2}$$

The largest eigenvalue provides information on the posterior distribution of the magnitude of the first principal component of each development year, decomposed by accident year. That is, we can quantify in the PIC model, by accident year, the proportion of residual variation in the log payments for accident year i currently unexplained by the development factors $\Phi_{0:J-i}$, which were jointly estimated in the PIC model and assumed constant accross each accident year (i.e. constant per development year) for parsimony. We can also repeat this for the incurred loss data. Suppose that a principal component analysis is performed, decomposing the variation in the payment and incurred data for each accident year i with respect to the variation unexplained by the development factors in the PIC model. Then, up to proportionality, the distribution of the eigenvalues corresponds to the proportion of contribution from the leading eigenvector (principal component). When this is coupled with the fact that we can also easily obtain samples from the marginal posterior distribution of the leading eigenvector of the covariance matrix for the i-th accident year's payment of incurred loss data in the PIC model, then we get complete information per accident year on the ability of the development factors in the PIC model to explain variation in the observed loss data. Table 4.9 summarises the results for the average PCA weight (largest eigenvalue) and average posterior eigenvector.

Tokuda et al. (2011) develops a framework which formalizes an approach to the summary of dependence structures. For the running example of results that we present for distributions $p\left(\Sigma_4^P|\boldsymbol{P},\boldsymbol{I}\right)$ and $p\left(\Sigma_4^I|\boldsymbol{P},\boldsymbol{I}\right)$, under such an approach the third and fourth layers of summary are presented in Figure 4.3. Figure 4.3 represents the Heatmaps for the block diagonal covariance matricesΣ_4^P (2×2 sub-plot 1) and Σ_4^I (2×2 sub-plot 2). These are obtained using samples from the Riemannian Manifold Adaptive Metropolis sampler. Samples from the Posterior distribution of the telescoping diagonal block size form covariance matrix structures of the Gaussian copula under the hierarchical Bayesian model which has the joint covariance between all observed payment and incurred loss data under the dependent development years. Each set of 4×4 panels, starting from the top, summarizes the posterior distributions for the covariance matrices for $s \in \{P, I\}$ according to: ***Top Left Panel***: contour map of posterior samples $\log\left[\Sigma_4^s\right]_{1,1}$ vs $\log\left[\Sigma_4^s\right]_{5,5}$. ***Top Right Panel***: contour map of posterior samples $\log\left[\Sigma_4^s\right]_{1,1}$ vs $\left[\Sigma_4^s\right]_{1,5}$. ***Bottom Left Panel***: kernel density estimator of the posterior distribution of the trace of the covariance matrix using samples $\left\{\log \mathrm{tr}\left(\Sigma_4^s\right)\right\}$. ***Bottom Right Panel***: scatter plot of posterior samples of the first, second and third largest eigenvalues scaled by total of the eigen valuse - (PCA weights - for linear combinations of the development factors when explaining variation in observed payment and incurred data for a given accident year). This involves the presentation of contour maps of these marginal posteriors that are constructed using adaptive MCMC samples of these matrices.

In Figure 4.4, the development factors for payment and incurred data marginal posterior distributions are presented along with the posteriors of the hyperparameters for the Gaussian Copula based PIC models (Model III). The Boxplot summaries of the marginal posterior distributions obtained using samples from the Riemannian Manifold Adaptive Metropolis sampler. Samples from the Posterior distribution under a telescoping diagonal block size form covariance matrix structures of the Gaussian copula under the hierarchical Bayesian model which has the joint covariance between all observed payment and incurred loss data under the dependent development years. ***Top Left Panel***: box plots of marginal posterior distributions for $p\left(\Phi_i|\boldsymbol{P},\boldsymbol{I}\right)$. ***Top Right Panel***: box plots of marginal posterior distributions for $p\left(\Psi_i|\boldsymbol{P},\boldsymbol{I}\right)$. ***Bottom Left Panel***: box plots of marginal posterior distributions for $p\left(s_i|\boldsymbol{P},\boldsymbol{I}\right)$. ***Bottom Right Panel***: box plots of marginal posterior distributions for $p\left(t_i|\boldsymbol{P},\boldsymbol{I}\right)$. Finally, we also compare the estimated posterior marginal distributions of the development factors for the payment and incurred loss triangles for the models: payment only model; the incurred only model; the Gaussian Copula (Model III) dependent model; the PIC [Full] independent model and the PIC [Partial] independent model of Merz and Wuthrich. (2010). The results of this comparison include the posterior mean estimates of $\mathbb{E}\left[\Phi_i|\boldsymbol{P},\boldsymbol{I}\right]$ and $\mathbb{E}\left[\Psi_i|\boldsymbol{P},\boldsymbol{I}\right]$, for all $i \in \{0, 1, \ldots, J\}$ and the posterior quantiles for left and right tails as measured by the fifth and ninety-fifth percentiles, given in Table 4.1. *We note that the results in this section for the Gaussian copula models are obtained using the log ratio observational data and the restults for the Mixture Archimedian copula model are more conveniently obtained using the log observations (not ratio data).*

It is also worth noting other approaches that can be adopted in the case of the Gaussian copula model. One could also included a data-augmentation stage in the analysis as was utilised in the Mixture Archimedian copula example. In addition, the covariance matrices could have been specified under different structures with more or less parsimony. The examples utilised in this section were those which provided a reasonable trade-off between parsimonious model specification, while allowing a meaningful decomposition of the results.

The results of the comparison between the Gaussian copula PIC model and the independent PIC model illustrated that whilst the posterior marginal mean development factor estimates are not affected by the dependence feature included, the marginal posterior shape is affected. This is reflected by the comparison of the posterior confidence intervals for the Gaussian copula PIC model when compared to the payment or incurred individual models where there is a significant difference present in the shapes of the marginal posterior. It is expected that this will have implications for the estimation of reserves using these different will be quantified in the next section.

4.6.1.2. *Data-Augmented hierarchical Bayesian Mixture-Archimedian Copula PIC (Model IV).* This section presents the estimation results for the mixture of Clayton and Gumbel Copula based PIC models (Model IV) on the real data are presented in this section. Figure 4.5 presents a summary of the mixture copula dependence structure obtained from posterior samples of the copula parameters under the hierarchical Bayesian model. More specifically, copula Dependence Parameter Posterior distributions estimated under the Data-Augmented Mixture Copula PIC Model IV. A mixture of Archimedean copula models is considered, with Clayton and Gumbel copula choices, allow for possible asymmetry in the tail dependence over development years.We chose uniformative uniform priors $U[0, 20]$ for the copula parameters. ***Top Left Panel***: Contour map of posterior estimated mixture copula dependence distribution between development years over paid and incurred loss data, with homogeneous dependence assumptions over accident years (estimated from posterior mean of ρ_C^{MMSE} and ρ_G^{MMSE}. ***Top Right Panel***: Surface plot of posterior estimated mixture copula dependence distribution between development years over paid and incurred loss data, with homogeneous dependence assumptions over accident years (estimated from posterior mean of ρ_C^{MMSE} and ρ_G^{MMSE}. ***Bottom Left Panel***: Scatter plot of posterior samples used to estimate Kendall's tau rank correlation versus copula parameter for the Clayton mixture component. ***Bottom Right Panel***: Scatter plot of posterior samples used to estimate Kendall's tau rank correlation versus copula parameter for the Gumbel mixture component. The results in this section are

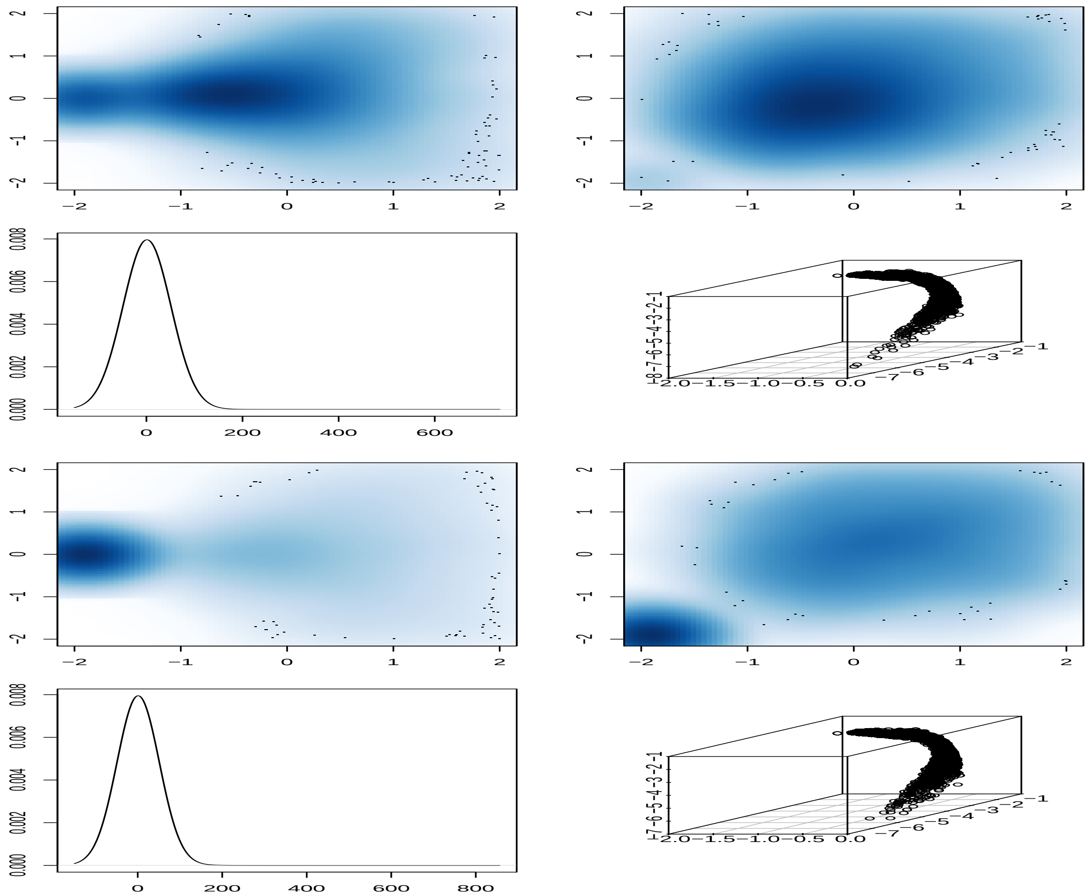


FIGURE 4.3. Heatmaps for the block diagonal covariance matrices

obtained using the log observational data, not ratio data. The figures summarise succinctly the estimated posterior dependence structure for the hierarchical Bayesian mixture Copula model, through plots of the dependence structure as captured by the estimatd mixture copula distribution, the scatter plots of copula parameter for the lower tail and rank correlation (Kendall's tau) and the upper tail copula parameter versus rank correlation. These results clearly demonstrate posterior evidence for non-trivial tail dependence features in the payment and incurred data, as well as potential for asymmetry in the upper and lower tail dependence. Note, uniformative prior choices were made on the copula parameters with uniform priors over $[0, 50]$ and $[1, 50]$ respectively, indicating these estimated copula parameters are data driven results.

Figure 4.6 presents the development factors for payment and incurred data marginal posterior distributions along with the hyperparameter marginal posteriors for the Data-Augmented Mixture Copula based PIC models (Model IV). It presents the boxplots of the marginal posterior distributions of the development factors and hyperparameters. ***Top Left Panel***: box plots of marginal posterior distributions for $p\left(\Phi_i|\boldsymbol{P},\boldsymbol{I}\right)$. ***Top Right Panel***: box plots of marginal posterior distributions for $p\left(\Psi_i|\boldsymbol{P},\boldsymbol{I}\right)$. ***Bottom Left Panel***: box plots of marginal posterior distributions for $p\left(s_i|\boldsymbol{P},\boldsymbol{I}\right)$. ***Bottom Right Panel***: box plots of marginal posterior distributions for $p\left(t_i|\boldsymbol{P},\boldsymbol{I}\right)$.

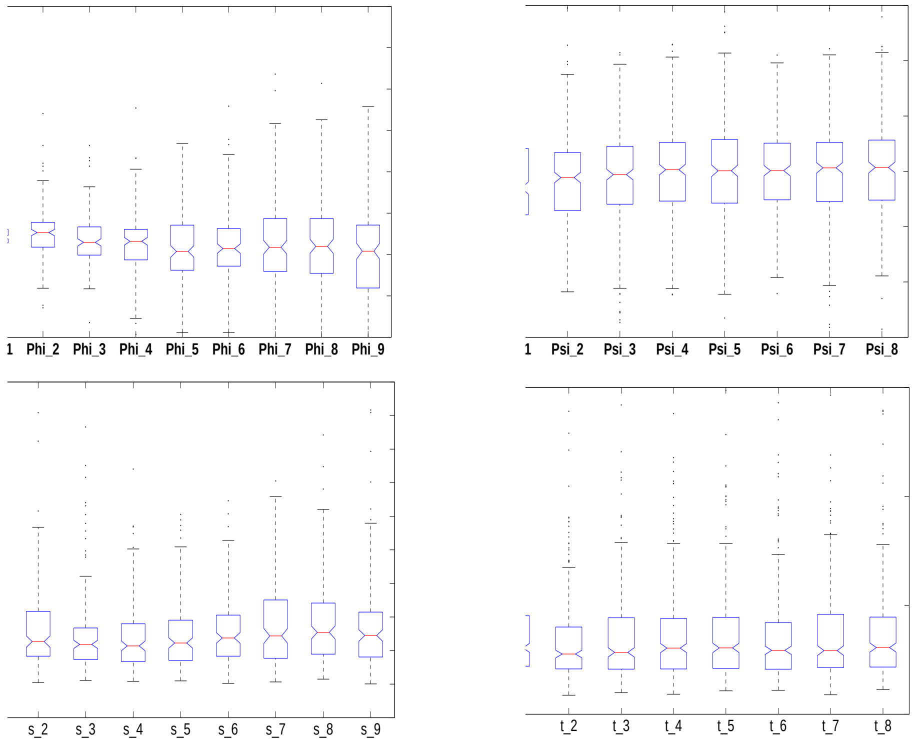


FIGURE 4.4. Boxplot of the marginal posterior distributions

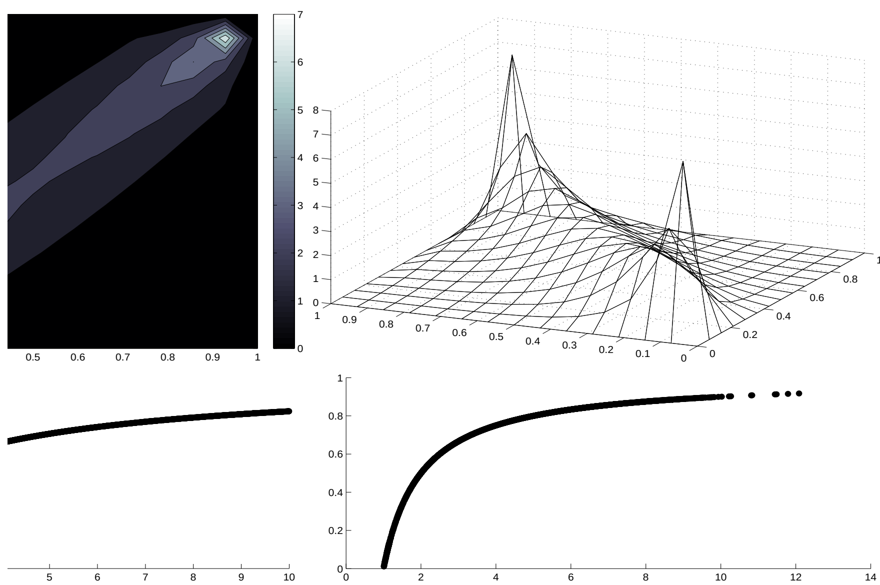


FIGURE 4.5. Copula Dependence Parameter Posterior distributions

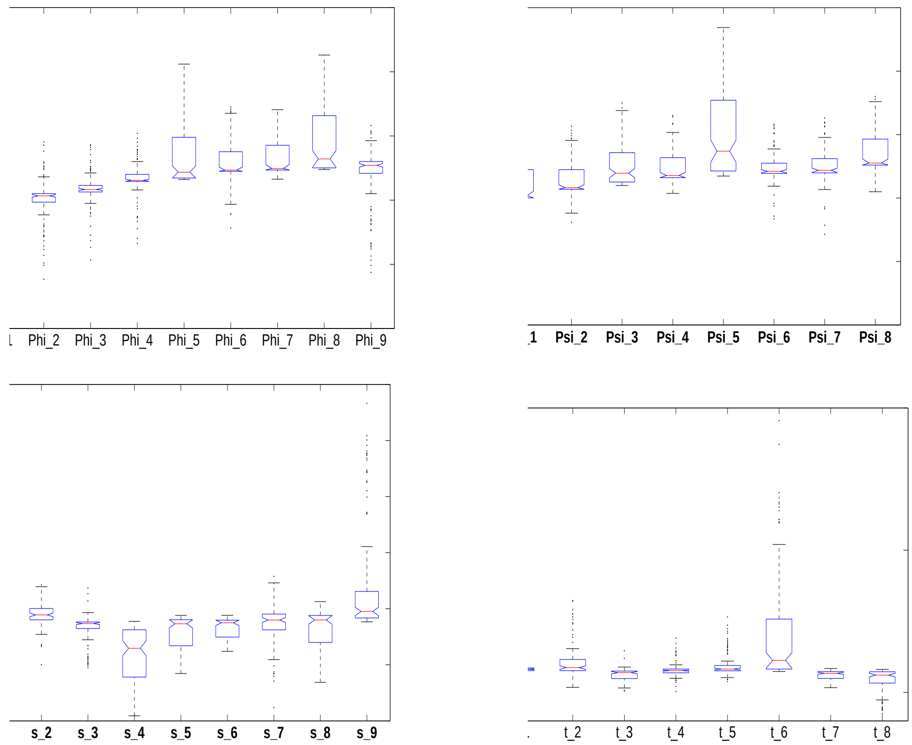


FIGURE 4.6. Boxplots of the marginal posterior distributions of the development factors and hyperparameters.

4.7. Comparison of PIC reserving with Gaussian Copula PIC and Mixture Archimedian Copula PIC Models

This section discuss the effect of modelling the dependence structures on the reserving estimates. First we note two important details in calculating the reserves. We need to be able to draw samples from the predictive distributions for the payment and incurred data given below, for each accident year i, using

$$p\left(P_{i,J}|\boldsymbol{P},\boldsymbol{I}\right)=\int p\left(P_{i,J}|P_{i,1:J-i},\boldsymbol{\Theta}\right)p\left(\boldsymbol{\Theta}|\boldsymbol{P},\boldsymbol{I}\right)d\boldsymbol{\Theta}\ \text{ and }\ p\left(I_{i,J}|\boldsymbol{P},\boldsymbol{I}\right)=\int p\left(I_{i,J}|I_{i,1:J-i},\boldsymbol{\Theta}\right)p\left(\boldsymbol{\Theta}|\boldsymbol{P},\boldsymbol{I}\right)d\boldsymbol{\Theta}$$

In general it is not possible to solve these integrals analytically. However, for the Gaussian copula models developed in this chapter, under the results in Lemma 4.3.3, one adopt the results of Merz and Wuthrich. (2010)[Theorem 2.4] to obtain analytic Gaussian predictive distributions. Alternatively, the predicitive distributions can be estimated as described in **?**, Section 3.3. Although the results in Table 4.1 demonstrate that the incorporation of the dependence structures does not significantly alter the posterior mean of the development factors for the payment and incurred loss data, it is clearly possible for the predictive distribution to be altered, since the shape of the posterior distribution is altered by the dependence features. Second, regarding the hierarchical mixture Archimedian copula model, it does not admit an analytic solution for the predictive distribution. This does not matter if a data augmentation stage is set up in the joint posterior distribution to sample cumulative payments, since then we can use the MCMC sampler output for the ultimate cumulative payment and incurred losses in each accident year.

Finally, we also note that a simple Monte Carlo based approximation for the ultimate claim can be constructed. Take the samples from the MCMC output for the PIC model of interest (sampled from the complete PIC model with dependence features present) and then utilise these samples to construct a Laplace approximation to the predictive observation distribution for example $p\left(P_{i,J}|P_{i,1:J-i},\boldsymbol{\Theta}\right)$ which involves a normal approximation around the MAP or locally around each Monte carlo sample for the development factors, with precision given by the sampled observation covariance structure. Though this is not required, as we have shown for the Gaussian copula models independence models, it may be useful for alternative copula based models with simple data-augmentation approaches. In addition a second alternative would be to utilise in the predictive distribution the marginal distributions.

Figure 4.7 presents the log posterior predictive distribution for the ultimate total claim given by the predictive distribution for the log of the cumulative payment over each accident year $\sum_{i=0}^{J} P_{i,J}$ for the full Bayesain models which incorporate priors on observation error, development factors and hyperpriors for precision of the development factors. Ultimate Bayesian predictive distributions for log payment data from the payment only predictive distribution, the Full Independent PIC model, and the hierarchical PIC Mixture Copula model via Data Augmentation predictive distribution. ***Left Panel***: Posterior predictive distribution box plots from samples. ***Right Panel***: Kernel density estimates of the predictive distributions. We see that all three models are in good agreement with each other with the dependence parameters affecting the variance and tail behaviour of the distributions.

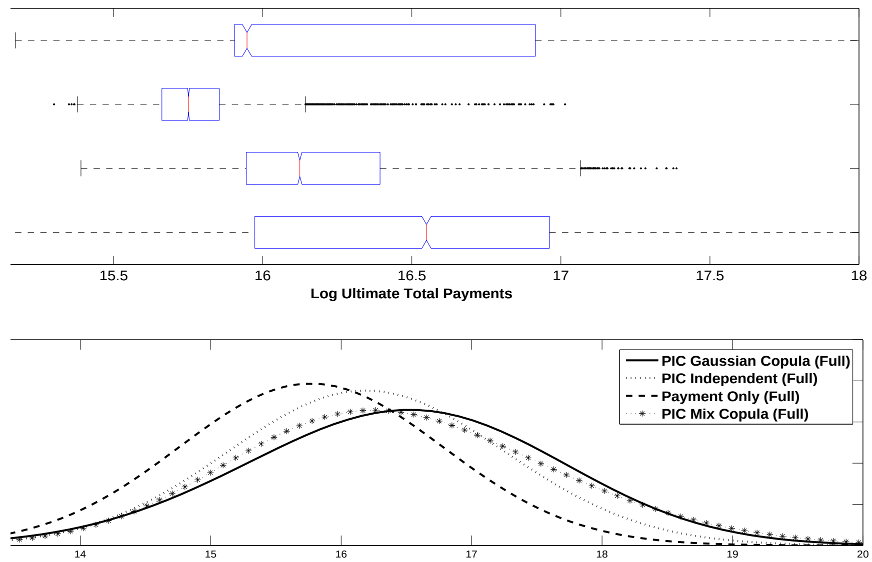


FIGURE 4.7. Boxplots of the predictive distributions obtained from the MCMC samples

Next we consider the distributions of the outstanding loss liabilities estimated using the S samples from the MCMC obtained for the posterior PIC model. We denoted these by random variables $\left\{R(\boldsymbol{P},\boldsymbol{I})^{(s)}\right\}_{s=1:S}$ where $R(\boldsymbol{P},\boldsymbol{I})^{(s)} = P_{i,J} - P_{i,J-i}$ and depending on whether payment, incurred, or both data is present we denoted by $R(\boldsymbol{P})^{(s)}$, $R(\boldsymbol{I})^{(s)}$ and $R(\boldsymbol{P},\boldsymbol{I})^{(s)}$ respectively. Figure 4.8 presents the MCMC estimated claims reserve marginal posterior predictive distributions for each accident year per model developed. It is the boxplots of log ultimate Bayesian predictive reserve distributions for payment data per accident year, compared to (Partial) PIC Independent posterior mean estimates from Merz and Wuthrich. (2010) (karge unfilled black circles). **Top Row:** the (Full) hierarchical PIC Mixture Copula model via Data Augmentation; **Second Row from Top:** the (Full) hierarchical PIC Gaussian Copula model; **Third Row from Top:** the (Full) Independent PIC model; **Bottom Row:** the (Full) payment Only model.

We compared our results to those obtained in Merz and Wuthrich. (2010) and find good agreement between the mean reserve per accident year and each proposed model. In addition, we note the possible differences between the distributions can be attributed to the utilisation of the full versus partial hierarchical Bayesian models in this paper and the different dependence structures considered. Additionally, we note that further analysis on comparisons to existing models in the

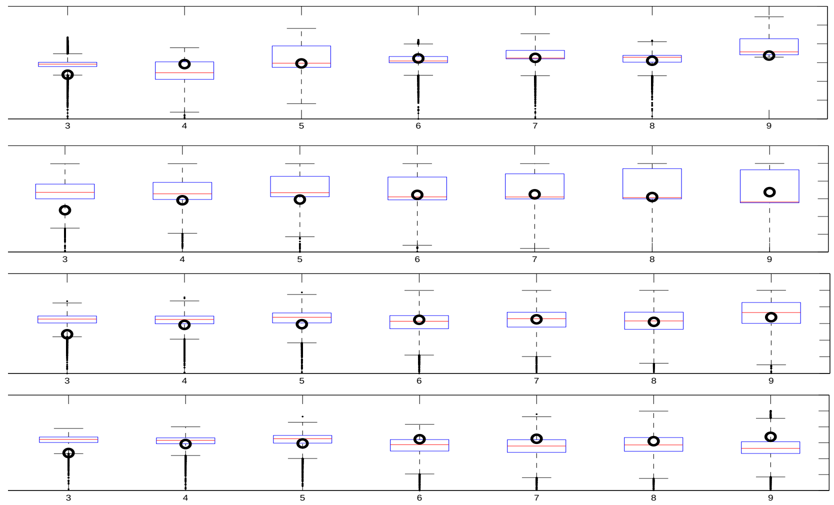


FIGURE 4.8. Boxplots of log ultimate Bayesian predictive reserve distributions for payment data per accident year

literature can be obained for the models of **?**, Dahms (2008) and Quarg and Mack (2004) for this data analysis in Merz and Wuthrich. (2010) [Table 4] and in the spreadsheet provided by Professor Mario Wuethrich at URL[1].

4.8. Conclusions

This chapter extends the class of PIC models to combine the two different channels of information as proposed in Merz and Wuthrich. (2010) by introducing several novel statistical models for the dependence features present within and between the payment and incurred loss data. This allows us to obtain a unified ultimate loss prediction which incorporates the potential for general dependence features. To achieve this we developed full hierarchical Bayesian models which incorporate several different potential forms of dependence, including generalized covariance matrix structure priors based on inverse Wishart distributions and conditional Bayesian conjugacy in the PIC independent log-normal model. This forms a general class of Gaussian copula models which extends the approach of Happ and Wuthrich (2011).

Second, we develop a class of hierarchical mixture Archimedian copula models to capture potential for tail dependence in the payment and incurred loss data, again developing and demonstrating how to appropriately construct a full Bayesian model incorporating hyperpriors. In this regard, we also develop a class of models in which data-augmentation is incorporated to both overcome challenging marginal likelihood evaluations required for the MCMC methodology to sample from

[1] URL:`http://www.math.ethz.ch/~wueth/claims_reserving3.html`

the PIC Bayesian models. This had the additional feature that it also allowed for joint Bayesian inference of the reserves as part of the posterior inference.

Finally, to perform inference on these approaches we developed an adaptive Markov chain Monte Carlo sampling methodology incorporating novel adaptive Riemann-manifold proposals restricted to manifold spaces (postive definite symmetric matrices) to sample efficiently the covariance matrices in the posterior marginal for the Gaussian copula dependence. We make these advanced MCMC accessible to the actuarial audience to address challenging Bayesian inference problems in Claims Reserving modelling.

The consequence of these models for actuaries is that a new extended suite of flexible dependence structures have been incorporated into the recently proposed PIC models. These can now be extended and compared to existing chain ladder methods. We perform an analysis on real payment and incurred loss data discussed in Merz and Wuthrich. (2010) and compare our models with the analysis for the independent PIC model (partial) and the (full) Bayesian PIC model as well as several different dependent models and the payment only model. Furthermore, we provide reference on further comparisons to the alternative models of **?**, Dahms (2008) and Quarg and Mack (2004) for this data.

4.9. Acknowledgement

We would like to thank Prof. Mario Wuethrich of ETH for his discussions on this topic and for introducing us to the family of models that we have explored. The research of Robert Kohn was partially supported by ARC Discovery Grant DP0667069.

Posterior Covariance Matrix for payments and incurred Loss Gaussian Copula

Sub-Block	Ave. $\widehat{\lambda}_i^{(s)}$	Std.Dev $\widehat{\lambda}_i^{(s)}$	$[Q0.05; Q0.95]$ for $\widehat{\lambda}_i^{(s)}$	Ave. Principal Eigen Vector $\widehat{\boldsymbol{v}}$
Σ_0^P	2.52	13.15	[0.15;11.09]	[0.10,-0.12,-0.07,-0.03,-0.05,-0.03,0.01,-0.06,-0.03]
Σ_1^P	1.97	13.92	[0.15;8.30]	[0.05,-0.08,-0.04,-0.02,-0.01,-0.01,-0.02,-0.01,-0.01]
Σ_2^P	0.94	8.48	[0.14;3.19]	[0.06,-0.10,-0.06,-0.02,-0.02,-0.01,-1.1e-3,-0.01]
Σ_3^P	0.75	6.38	[0.14;1.92]	[0.08,-0.12,-0.05,-0.03,-0.01,-6.9e-5,-2.3e-3]
Σ_4^P	0.76	6.81	[0.13;0.25]	[0.12,-0.13,-0.06,-0.03,-0.01,0.01]
Σ_5^P	0.70	5.93	[0.12;0.23]	[0.14,-0.15,-0.08,-0.01,0.01]
Σ_6^P	1.11	9.90	[0.12;0.24]	[0.21,-0.20,-0.07,0.03]
Σ_7^P	2.16	18.07	[0.10;0.26]	[0.27,-0.25,0.07]
Σ_8^P	5.44	34.67	[0.08;20.92]	[-0.47,0.43]
Σ_9^P	1.95	10.28	[0.06;11.56]	Not Applicable
Σ_0^I	1.69	4.78	[0.13;6.57]	[0.10,-0.12,-0.07,-0.03,-0.05,-0.03,0.01,-0.06,-0.03]
Σ_1^I	1.08	3.66	[0.13;4.93]	[0.03,-0.12,-0.05,-0.03,-0.01,2.4e-3,0.01,0.01]
Σ_2^I	0.80	3.26	[0.12;3.76]	[0.09,-0.12,-0.06,-0.02,-0.01,3.2e-3,-0.01]
Σ_3^I	0.66	3.15	[0.11;3.27]	[0.10,-0.12,-0.07,-0.03,-0.02,1.5e-3]
Σ_4^I	0.65	3.97	[0.10;3.23]	[0.15,-0.15,-0.07,-0.02,0.02,]
Σ_5^I	1.00	5.02	[0.09;5.62]	[6.6e-12,-6.6e-12,-2.9e-12,-1.0e-12,8.5e-13]
Σ_6^I	1.15	7.52	[0.08;6.31]	[0.31,-0.24,0.07]
Σ_7^I	5.26	25.29	[0.06;32.87]	[-0.50,0.42]
Σ_8^I	1.03	3.82	[0.04;6.30]	Not Applicable

NOTE: **(Full)** corresponds to PIC models with results for the FULL hierarchical Bayesian PIC model with priors on development factors, observation variances and hyperpriors on precisions on development factors. The PIC Independent **(Partial)** of Merz and Wuthrich. (2010) are the Bayesian posterior results in which σ_m and τ_n are assumed known. In addition, the PIC Mixture Copula model has posterior development factors on the scale of log cumulative payment data (not ratio data), so the reported posterior mean development factors are for the cumulative payment marginal posterior means (log scale).

Posterior Marginal Distributions for Development Factors

Factor	PIC Gaussian Copula (Full)		PIC Independent (Full)		payment or incurred only (Full)		Merz and Wuthr (2010) PIC Indep dent (Partial)
	Post.Ave.	$[Q0.05; Q0.95]$	Post.Ave.	$[Q0.05; Q0.95]$	Post.Ave.	$[Q0.05; Q0.95]$	Post.Ave.
Φ_0	13.79	[13.55;14.03]	14.51	[13.19;15.01]	13.77	[13.68;13.86]	13.78
Φ_1	0.21	[-0.16;0.58]	0.18	[0.05;0.29]	0.20	[0.12;0.27]	0.22
Φ_2	0.25	[-0.25;0.77]	0.22	[0.08;0.34]	0.23	[0.14;0.31]	0.24
Φ_3	0.18	[-0.44;0.81]	0.17	[0.04;0.30]	0.15	[0.06;0.24]	0.17
Φ_4	0.15	[-0.55;0.86]	0.16	[0.02;0.30]	0.13	[0.04;0.23]	0.16
Φ_5	0.13	[-0.63;0.91]	0.15	[1.9e-3;0.30]	0.12	[0.01;0.22]	0.14
Φ_6	0.10	[-0.71;0.92]	0.12	[-0.04;0.30]	0.08	[-0.04;0.20]	0.11
Φ_7	0.07	[-0.79;0.93]	0.13	[-0.05;0.33]	0.05	[-0.09;0.19]	0.07
Φ_8	0.08	[-0.81;0.97]	0.11	[-0.09;0.32]	0.05	[-0.12;0.22]	0.05
Φ_9	0.04	[-0.88;0.98]	0.10	[-0.04;0.52]	0.02	[-0.19;0.24]	0.08
Ψ_0	0.51	[-0.84;1.85]	0.45	[0.31;0.56]	0.52	[0.38;0.64]	0.50
Ψ_1	-0.15	[-1.50;1.20]	-0.08	[-0.11;0.12]	0.01	[-0.11;0.12]	-0.15
Ψ_2	-0.13	[-1.49;1.23]	-0.09	[-0.15;0.20]	0.01	[-0.12;0.12]	-0.14
Ψ_3	-3.7e-2	[-1.39;1.34]	0.01	[-0.05;0.21]	0.01	[-0.13;0.13]	-0.04
Ψ_4	-1.7e-2	[-1.39;1.36]	-0.01	[-0.06;0.23]	-0.01	[-0.15;0.14]	-0.02
Ψ_5	-7.1e-3	[-1.39;1.38]	0.02	[-0.06;0.21]	-0.06	[-0.17;0.15]	-0.02
Ψ_6	-7.3e-3	[-1.40;1.39]	-0.02	[-0.05;0.30]	-0.01	[-0.19;0.16]	-0.01
Ψ_7	-2.4e-3	[-1.40;1.39]	0.02	[-0.05;0.34]	-0.06	[-0.40;0.22]	-0.01
Ψ_8	-2.0e-4	[-1.40;1.40]	-0.01	[-0.02;0.52]	-0.13	[-0.52;0.25]	-0.01

TABLE 4.1. Results for Bayesian PIC model

CHAPTER 5

Summary

This section summaries how each chapter in this thesis forms some of the main building blocks in claims reserving.

5.1. Overview

In claims reserving or the valuation of insurance liabilities, the aim is to estimate the future claims experience which is to be expected on the business written to date by the insurer. There are several main building blocks involved.

As a first step, the estimation of central estimate represents the main part of insurance liabilities. It is therefore important that central estimate are valued in a realistic and appropriate manner. Chapter two of this thesis focuses on developing models for estimating central estimate for long tail insurance business classes in which there is large degree of uncertainty in setting reserve. A Bayesian approach is presented to model loss reserving data using the flexible GB2 distribution and dynamic mean models. The proposed GB2 distribution provides a flexible probability density function, which nests various distributions with light and heavy tails, to facilitate accurate loss reserving in insurance applications. Furthermore, we also extend the mean functions to include the state space and threshold models provides a dynamic approach to allow for irregular claims behaviors and legislative change which may occur during the claims settlement period. Apart from aggregated data, we have also considered models for individual claims data. The mixture of GB2 distributions is proposed as a mean of modeling the unobserved heterogeneity which arises from the incidence of very large claims in the loss reserving data. It is demonstrated through both simulation study and forecasting that model parameters are estimated in high accuracy, and the proposed models outperform the transitional models.

The second lays of insurance liabilities applied on top of central estimate is the risk margin. The risk margin is an allowance for the variability of claims experience, and together with the central estimate to produce a reasonable valuation of the insurance liabilities at a certain probability of sufficiency level. In chapter three, we propose the use of quantile regression models to derive risk margin and evaluate capital. We demonstrate quantile regression is capable of providing an accurate estimation of risk margin and overview of implied capital based on the historical volatility of a general insurer's loss portfolio. Two modeling frameworks are considered based around parametric and nonparametric quantile regression models which we contrast specifically in this insurance setting.

In the parametric quantile regression setting, we consider several models including the flexible generalized beta distribution family, asymmetric Laplace distribution and power Pareto distribution. These models are developed under a Bayesian regression framework via MCMC sampling strategies. In the nonparametric quantile regression models, we adopted AL distribution as a proxy and together with the parametric AL model in which we expressed the solution as a scale mixture of uniform distributions to facilitate implementation. Then we extend the best performed model, which is the AL model with ANOVA mean and variance functions, to estimate risk margin on two real loss reserve data sets. The generalized AL model with a dynamic shape parameter p provides us a mathematically consistent way of estimating risk margin. Overall, the results of our studies indicate that this new risk margins framework offers considerable potential benefits for reserving purpose.

From the perspective of good claims reserving practice, incorporating multiple sources of information and quantifying the predictive variability is of more interest than forecasting outstanding claims itself. The fact that we have two sources of information, namely the payment amount and incurred cost, allows us to study the dependence among the two source of data in the determination of a sufficient reserve and its associated variation. Examining two loss triangles jointly improves the accuracy in the prediction of losses by borrowing information from each other. In chapter four, we consider the class of recently developed family of hierarchical Bayesian Paid-Incurred-Claims models that combine claims payments and incurred losses information into a coherent reserving methodology. In particular, we extend the independent log-normal model by incorporating different dependence structures using a Data-Augmented mixture Copula Paid-Incurred claims model.

The utility and influence of joint modelling of a run-off triangle with Paid data on the one hand and Incurred claims on the other hand is demonstrated via both of independent models and dependent models. We investigate two proposals for a dependence structure between Paid and Incurred triangles, namely the lag-year telescoping block diagonal Gaussian Copula PIC data model incorporating conjugacy via transformation and the data-augmented mixture Archimedean copula dependent PIC data model. They are implemented via a class of adaptive Markov chain Monte Carlo (MCMC) sampling algorithms. It is shown through our study that the proposed models which incorporate two sources of information improve estimate accuracy in claims reserving.

5.2. Executive summary

This thesis proposes a Bayesian dynamic reserving framework. The GB2 distribution and its mixture representation with adynamic mean and variance are considered for the estimation of central estimate for long tail insurance business class where tail behaviors can be largely different. We applied the models to two real insurance data sets, one being aggregated claims triangle and the other one being the individual claims data. It is shown through both simulation study and forecasting that model parameters are estimated in high accuracy, and the proposed models outperform the transitional models including Chain ladder and Gamma models.

Under this dynamic reserving framework, risk quantities, such as risk margin and VaR are estimated by quantile regression models. We compare the performance of both parametric and non-parametric quantile regression models on two real insurance data sets. In the parametric framework, we built five models, namely AL, PP, GB2, GG and gamma. The AL model provides the best fit. We also investigate three different regression structures, namely ANCOVA, ANOVA and Poisson-Tweedie regression. The ANOVA model performs the best in our empirical data study. It is also demonstrated that the AL distribution with a dynamic shape parameter p provides a mathematically mechanism to estimate risk margin scientifically.

Our proposed dynamic reserving models also consider modelling two sources of reserving data sets jointly. We extend the class of PIC models as proposed in Merz and Wuthrich. (2010) by introducing several novel statistical models for the dependence features present within and between the payment and incurred loss data. This allows us to obtain a unified ultimate loss prediction which incorporates the potential for general dependence features.

5.3. Further research

In spite of the substantial development of Bayeisan loss reserve models in this thesis, there are still many unattended areas that are worthy of further consideration.

5.3.1. Extension to multivariate quantile regression. As many claim processes are jointly correlated over time, it may be more efficiently modeled via a multivariate structure. Since there are multiple sources of reserve data, the presence of dependency is inevitable. Depending on the nature of the data, the dependence structure can be vastly different. Multivariate regression quantile model can be considered to directly study the degree of tail interdependence among different types of data as well as between accident and development years to derive risk measures, such as risk margin and VaR. To capture these dependence structures explicitly, we can extend our models to consider the joint dependence structure of the multivariate random vector conditional on the covariate structure though a copula or other alternative approach.

5.3.2. Estimation of diversification effect. In general insurance reserving practice, actuaries quantify the unpaid losses and associated uncertainty for each line of business separately, and then aggregate the reserves from multiple lines to determine the company-level reserve. The diversification effect arises due to the association among different lines of business. It usually has a critical implication for the company-level reserve, and hence it is critical to estimate it scientifically. Multivariate quantile regression can be extend to model different lines of business within a company for setting company-level risk margin and reserve. Depending on the nature of the different lines of businesses, the various sources of dependency may appear. The correlations correlations might across development years as they develop over time or among accident years when natural disaster happens and effect multiple lines of business. Some People have focussed on correlations over calendar years based on the assumption that inflationary trends as a common unknown factor inducing correlation. These correlations can be modeled by different considering different dependent structures in the multivariate quantile regression model.

5.3.3. Comparison in methodologies of inference. In this thesis, we focus on the extensions of loss reserve models to account for different risk measures under the Bayesian framework. Another area that is worthwhile to study is the inference methodology. Though using efficient MCMC strategies improve the computation time, the sampling time could still be massive if the data size is large and the model is highly complicated. As an alternative would be the maximum likelihood method in the frequentist approach and the Expectation-maximization algorithm method which is an iterative method for finding maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables. For modelling individual claims data from large insurance portfolio, the maximum likelihood method and Expectation-maximization algorithm method can be considered.

1. Appendix A

Lemma 1.1 (Properties of Matrix-Variate Gaussian Distribution). *A $p \times n$ random matrix X is said to have a matrix variate Gaussian distribution with $p \times n$ mean matrix M and covariance matrix $\Sigma \otimes \Psi$ where Σ and Ψ are in the spaces of symmetric positive definite matrices given by $\Sigma \in \mathbb{SD}^{+}(\mathbb{R}^{p})$ and $\Psi \in \mathbb{SD}^{+}(\mathbb{R}^{n})$ if the $pn \times 1$ dimensional random vector $Vec\,(X')$ has a multivariate normal distribution $Vec\,(X') \sim N\,(Vec(M'), \Sigma \otimes \Psi)$. Furthermore, if X is distributed according to matrix-variate Gaussian distribution $X \sim N_{p,n}\,(M, \Sigma \otimes \Psi)$ then the density is given by*

$$f_X^{MVN}\,(x; M, \Sigma, \Psi) = \frac{\exp\left(-\frac{1}{2}tr\left[\Sigma^{-1}\,(X-M)'\,\Psi^{-1}\,(X-M)\right]\right)}{(2\pi)^{np/2}\,|\Sigma|^{n/2}\,|\Psi|^{p/2}} \tag{.1}$$

In addition the following properties are satisfied for a matrix-variate Guassian (see Gupta and Nagar (2000) Chapter 2):

(1) *If $X \sim N_{p,n}\,(M, \Sigma \otimes \Psi)$, then $X' \sim N_{n,p}\,(M', \Psi \otimes \Sigma)$*

(2) *If $X \sim N_{p,n}\,(M, \Sigma \otimes \Psi)$, and partition X, M, Σ, and Ψ as*

$$X = \begin{bmatrix} X_{1r} \\ X_{2r} \end{bmatrix}, \text{ and } X = \begin{bmatrix} X_{1c} & X_{2c} \end{bmatrix} \tag{.2}$$

with X_{1r} the $(m \times n)$ sub-matrix, X_{2r} the $(p - m \times n)$ sub-matrix, X_{1c} the $(p \times t)$ sub-matrix and X_{2c} the $(p \times n - t)$ sub-matrix. With analogous partitions of the mean matrix M_{1r}, M_{2r}, M_{1c} and M_{2c} and covariance matrices

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \text{ and } \Psi = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}, \tag{.3}$$

with Σ_{11} the $(m \times m)$ sub-matrix, Σ_{12} the $(m \times p - m)$ sub-matrix, Σ_{22} the $(p - m \times p - m)$ sub-matrix, Ψ_{11} the $(t \times t)$ sub-matrix, Ψ_{22} the $(n - t \times n - t)$ sub-matrix. Then the following properties are true

$$\begin{aligned} X_{1r} &\sim N_{m,n}\,(M_{1r}, \Sigma_{11} \otimes \Psi) \text{ and } X_{1c} \sim N_{p,t}\,(M_{1c}, \Sigma \otimes \Psi_{11}) \\ X_{2r}|X_{1r} &\sim N_{p-m,n}\left(M_{2r} + \Sigma_{21}\Sigma_{11}^{-1}\,(X_{1r} - M_{1r})\,, \Sigma_{22.1} \otimes \Psi\right) \\ X_{2c}|X_{1c} &\sim N_{p,n-t}\left(M_{2c} + (X_{1c} - M_{1c})\,\Psi_{11}^{-1}\Psi_{12}, \Sigma \otimes \Psi_{22.1}\right) \end{aligned} \tag{.4}$$

with $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ and $\Psi_{22.1} = \Psi_{22} - \Psi_{21}\Psi_{11}^{-1}\Psi_{12}$.

Lemma 1.2 (Properties of Matrix Variate Wishart Distributions). *Suppose that the $(p \times p)$ positive definite matrix Σ has an inverse Wishart distribution, with positive definite $(p \times p)$ scale matrix Λ,* *degrees of freedom parameter $k > p-1$, and density*

$$f(\Sigma|\Lambda,k) = \frac{|\Lambda|^{m/2}\,|\Sigma|^{-(k+p+1)/2}\,e^{-\text{trace}(\Lambda\Sigma^{-1})/2}}{2^{kp/2}\Gamma_p(k/2)}, \tag{.5}$$

where $\Gamma_p(\cdot)$ is the multivariate gamma function. The mean and mode of this distribution are given respectively by

$$\mathbb{E}\left[\Sigma|\Lambda,k\right] = \frac{1}{k-p-1}\Lambda, \text{ and } m\left(\Sigma\right) = \frac{1}{k+p+1}\Lambda. \tag{.6}$$

Furthermore, the following marginal and conditional properties of the inverse Wishart distribution are relevant. Consider a partition of the matrices Λ and Ψ as

$$\Lambda = \begin{bmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Lambda_{22}\end{bmatrix},\ \Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix} \tag{.7}$$

with Λ_{ij} and Σ_{ij} denoting $p_i \times p_j$ matrices, then the following properties are satisfied (see Gupta and Nagar (2000)[Chapter 3, Section 3.4]):

(1) *The random sub-matrix Σ_{11} is independent of $\Sigma_{11}^{-1}\Sigma_{12}$*

(2) *The marginal distribution of any sub matrix on the diagonal of the matrix Σ is distributed as inverse Wishart. For example, the sub random matrix Σ_{11} is as inverse Wishart with $\Sigma_{11} \sim \mathcal{IW}\left(\Lambda_{11}, k-p_2\right)$;*

(3) *The marginal distribution of sub random matrix $\Sigma_{22\cdot 1}$ is inverse Wishart $\Sigma_{22\cdot 1} \sim \mathcal{IW}\left(\Lambda_{22\cdot 1}, k\right)$.*

In Lemma 1.3 below we present details for the matrix-variate Inverse Wishart distribution, see Gupta and Nagar (2000)[Chapter 3.4, Definition 3.4.1 and Theorem 3.4.1]

Lemma 1.3 (Properties of Matrix-Variate Inverse Wishart Distribution)**.** *A random $p \times p$ matrix $V = \Sigma^{-1}$ is distributed as Inverse Wishart, with degrees of freedom m and $p \times p$ parameter matrix Ψ, denoted $V \sim IW_p(m,\Psi)$ with density*

$$f(\Sigma|\Psi,m) = \frac{2^{-1/2(m-p-1)p}|\Psi|^{1/2(m-p-1)}}{\Gamma_p\left[1/2(m-p-1)\right]|V|^{1/2m}}\text{etr}\left(-1/2V^{-1}\Psi\right),\ V > 0, \Psi > 0, m > 2p. \tag{.8}$$

2. Appendix B

The family of Archimedean copula models has the following useful properties presented in Lemma 2.1.

Lemma 2.1. *Let C be an Archimedean copula with generator φ. Then according to* **?**, *Lemma 4.1.2 and Theorem 4.1.5, the following properties hold:*

(1) *C is an Archimedean copula if it can be reprsented by*

$$C(u,v) = \varphi^{[-1]}\left(\varphi(u) + \varphi(v)\right)$$

where φ is the generator of this copula and is a continous, strictly decreasing function from $[0,1]$ to $[0,\infty]$ such that $\varphi(1) = 0$ and $\varphi^{[-1]}$ is the pseudo inverse of φ.

(2) *C is symmetric, $C(u,v) = C(v,u)\ \forall (u,v) \in [0,1]\times[0,1]$*

(3) *C is associative, $C(C(u,v),w) = C(u, C(v,w))\ \forall (u,v,w) \in [0,1]^3$.*

(4) *If $c > 0$ is any constant, then $c\varphi$ is a generator of C*

(5) *According to Denuit et al. (2005, Definition 4.7.6), the extension of the Archimedean copula family to n-dimensions is achieved by considering the strictly monotone generator function φ such that $\varphi : (0,1] \to \mathbb{R}^+$ with $\varphi(1) = 0$, then the resulting Archimedean copula can be expressed as,*

$$C(u_1, u_2, \ldots, u_n) = \varphi^{-1}\left(\sum_{i=1}^{n} \varphi(u_i)\right).$$

The members of the Archimedean copula family utilised in this manuscript are given below in Lemma 2.2.

Lemma 2.2. *From the results in* **?**, *Section 4.3, Table 4.1 the distribution and density functions of the Clayton copula are given respectively as:*

$$C^C(u_1, ..., u_n) = \left(1 - n + \sum_{i=1}^{n} u_i^{-\rho^C}\right)^{-1/\rho^C}, \tag{.1}$$

$$c^C(u_1, ..., u_n) = \left(1 - n + \sum_{i=1}^{n} (u_i)^{-\rho^C}\right)^{-n-\frac{1}{\rho^C}} \prod_{i=1}^{n}\Big((u_i)^{-\rho^C-1}\big((i-1)\rho^C + 1\big)\Big), \tag{.2}$$

where $\rho^C \in [0,\infty)$ is the dependence parameter. The Clayton copula does not have upper tail dependence. Its lower tail dependence is $\lambda_L = 2^{-1/\rho^C}$. The distribution function of the Gumbel copula is

$$C^G(u_1, ..., u_d) = \exp\left(-\left[\sum_{i=1}^{d}\big(-\log(u_i)\big)^{\rho^G}\right]^{\frac{1}{\rho^G}}\right), \tag{.3}$$

where $\rho^G \in [1,\infty)$ is the dependence parameter. The Gumbel copula does not have lower tail dependence. The upper tail dependence of the Gumbel copula is $\lambda_U = 2 - 2^{1/\rho^G}$. The distribution function of the Frank copula is

$$C^F(u_1, ..., u_n) = \frac{1}{\rho}\ln\left(1 + \frac{\prod_{i=1}^{n}(e^{\rho^F u_i} - 1)}{(e^{\rho^F} - 1)^{n-1}}\right), \tag{.4}$$

where $\rho^F \in \mathbb{R}/\{0\}$ is the dependence parameter. The Frank copula does not have upper or lower tail dependence. We note that the density functions for Gumbel and Frank does not admit simple recursive expressions in terms of their density functions, but they can be obtained via partial differentiation

$$c(u_1, ..., u_n) = \frac{\partial^n}{\partial u_1, ...\partial u_n} C(u_1, ..., u_n). \tag{.5}$$

3102 ## 3. Appendix C

Proof *The proof of Lemma 4.4.1 requires one to demonstrate that the resulting distribution function*

$$\begin{aligned} \tilde{C}\left(u_1, u_2, \ldots, u_n\right) &= \int_{[0,u_1]\times[0,u_2]\times\cdots\times[0,u_n]} \tilde{c}\left(x_1, x_2, \ldots, x_n\right) dx_{1:n} \\ &= \sum_{i=1^m} w_i \int_{[0,u_1]\times[0,u_2]\times\cdots\times[0,u_n]} c_i\left(x_1, x_2, \ldots, x_n\right) dx_{1:n} \\ &= \sum_{i=1^m} w_i C_i\left(u_1, u_2, \ldots, u_n\right) \end{aligned}$$

satisfies the two conditions of a n-variate copula distribution given in [Definition 2.10.6] of **?**. *The first of these conditions requires that for every $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) \in [0,1]^n$, one can show that $\tilde{C}(\boldsymbol{u}) = 0$ if at least one coordinate of $\boldsymbol{u}$ is 0. Clearly since we have shown that $\tilde{C}(\boldsymbol{u}) = \sum_{i=1^m} w_i C_i(\boldsymbol{u})$ and given each member $C_i(u_1, u_2, \ldots, u_n) \in \mathcal{A}^n$ is define to be in the family of Archimedean copulas each of which therefore satisfies this condition for all such points $\boldsymbol{u}$, then it is trivial to see that the probability weighted sum of such points also satisfies this first condition. Secondly one must show that for every $\boldsymbol{a}$ and $\boldsymbol{b}$ in $[0,1]^n$, such that $\boldsymbol{a} \leq \boldsymbol{b}$ (i.e. $a_i < b_i\ \forall i \in \{1, 2, \ldots, n\}$) the following condition on the volume for copula $\tilde{C}$ is satisfied, $V_{\tilde{C}}([\boldsymbol{a}, \boldsymbol{b}]) \geq 0$. As in* **?** *we adopt the notation for the n-box, $[\boldsymbol{a}, \boldsymbol{b}]$, representing $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ and we define the n-box volume for copula distribution $\tilde{C}$ by [Definition 2.10.1, p.43] of* **?** *giving*

$$\begin{aligned} V_{\tilde{C}}([\boldsymbol{a}, \boldsymbol{b}]) &= \sum sgn(\boldsymbol{c}) \tilde{C}(\boldsymbol{c}) \\ &= \triangle_{a_1}^{b_1} \triangle_{a_2}^{b_2} \cdots \triangle_{a_n}^{b_n} \tilde{C}(\boldsymbol{c}) \end{aligned}$$

where the domain Dom$\tilde{C}$ of the mixture copula $\tilde{C}$ satisfies $[\boldsymbol{a}, \boldsymbol{b}] \subseteq$ Dom$\tilde{C}$. In addition we note that this sum is understood to be taken over all vertices $\boldsymbol{c}$ of n-box $[\boldsymbol{a}, \boldsymbol{b}]$ and sgn($\boldsymbol{c}$) $= 1$ if $c_k = a_k$ for an even number of k's or sgn($\boldsymbol{c}$) $= -1$ if $c_k = a_k$ for an odd number of k's. Equivalently, we consider $\triangle_{a_k}^{b_k} \tilde{C}(\boldsymbol{t}) = \tilde{C}(t_1, t_2, \ldots, t_{k-1}, b_k, t_{k+1}, \ldots, t_n) - \tilde{C}(t_1, t_2, \ldots, t_{k-1}, a_k, t_{k+1}, \ldots, t_n)$. In *the case of the mixture copula, we can expand the volume of the n-box* $[\boldsymbol{a}, \boldsymbol{b}]$ *as follows*

$$V_{\tilde{C}}([\boldsymbol{a}, \boldsymbol{b}]) = \sum sgn(\boldsymbol{c})\tilde{C}\left(\boldsymbol{c}\right) = \sum_{i=1}^{m} \sum w_i sgn(\boldsymbol{c}) C_i\left(\boldsymbol{c}\right) = \sum_{i=1}^{m} \sum w_i V_{C_i}([\boldsymbol{a}, \boldsymbol{b}])$$

hence we see that since each component $C_i\left(u_1, u_2, \ldots, u_n\right)$ *is a member of the set of Archimedean copula distributions* $\mathcal{A}^n$*, therefore for each component we have that* $V_{C_i}([\boldsymbol{a}, \boldsymbol{b}]) \geq 0$ *for all* $i \in \{1, 2, \ldots, m\}$.

$\square$

Bibliography

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.

Aiuppa, T. (1988). Evaluation of pearson curves as an approximation of the maximum probable annual aggregate loss. *Journal of Risk and Insurance*, 55:425Ű441.

Andrieu, C. and Moulines, E. (2006a). On the ergodicity properties of some adaptive mcmc algorithms. *Annals of Applied Probability*, 16:1462Ű1505.

Andrieu, C. and Moulines, É. (2006b). On the ergodicity properties of some adaptive mcmc algorithms. *The Annals of Applied Probability*, 16(3):1462–1505.

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive mcmc. *Statistics and Computing*, 18(4):343–373.

Andrieu, C. and Thoms., J. (2008). A tutorial on adaptive mcmc. *Statistics and Computing*, 18:343Ű373.

Atchadé, Y. and Rosenthal, J. (2005). On adaptive markov chain monte carlo algorithms. *Bernoulli*, 11(5):815–828.

Authority, A. P. R. (2012). Prudential standard gps 320. *Actuarial and Related Matters*, 5.

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis, second ed.* Springer-Verlag, New York.

Bingham, N., Goldie, C., and Teugels, J. (1989). *Regular variation*. Cambridge university press.

Björkwall, S., Hössjer, O., and Ohlsson, E. (2010). Bootstrapping the separation method in claims reserving. *ASTIN Bulletin*, 40:845–869.

Brown, A. (2012). Demystifying the risk margin: Theory, practice and regulation1. *the Staple Inn Actuarial Society*.

Cai, Y. (2010). Polynomial power-pareto quantile function models. *Extremes*, 13:291–314.

Celeux, G., Forbes, F., Robert, C., and Titterington, D. (2002). Deviance information criteria for missing data models. *The Review of Economics and Statistics*, 69:232Ű240.

Chan, J., Choy, S., and Makov, U. (2008). Robust bayesian analysis of loss reserves data using the generalized-t distribution. *Astin Bulletin*, 38:207–230.

Chen, Q., Gerlach, R., and Lu, Z. (2012). Bayesian value-at-risk and expected shortfall forecasting via the asymmetric laplace distribution. *Computational Statistics and Data Analysis*, 56:3498–3516.

Chib, S. and Greenberg, E. (1995). Understanding the metropolisŰhastings algorithm. *American Statistician*, 49:327Ű335.

Chiu, T., Leonard, T., and Tsui, K. (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association*, pages 198–210.

Cruz, M., Peters, G., and Shevchenko, P. (2014). *Advances in heavy tailed risk modeling: A Handbook of Operational Risk*. John Wiley and Sons.

Cummins, J., Dionne, G., M., and J.B., Pritchett, B. (1990). Application of the gb2 family of distributions in modelling insurance loss processes. *Insurance: Mathematics and Economics*, 9:257Ű272.

Cummins, J., Lewis, C., and Philips, R. (1999). *Pricing excess of loss reinsurance contracts against catastrophic loss. In: Froot, Kenneth (Ed.), The Financing of Catastrophe Risk*. University of Chicago Press, Chicago.

Cummins, J., McDonald, J., and Craig, M. (2007). Risk loss distributions and modelling the loss reserve pay-out tail. *Review of Applied Economics*, 3:1–23.

Dahms, R. (2008). A loss reserving method for incomplete claim data. *Bulletin Swiss Association of Actuaries*, pages 127–148.

Dahms, R., Merz, M., and Wüthrich, M. (2009). Claims development result for combined claims incurred and claims paid data. *Bulletin Francais dŠActuariat*, 9(18):5–39.

De Alba, E. (2002). Bayesian estimation of outstanding claim reserves. *North American Actuarial Journal*, 6:1–2.

De Jong, P. and Penzer, J. (2004). The arma model in state space form. *Statistics and Probability Letters*, 70:119Ű125.

De Jong, P. and Zehnwirth, B. (1983). Claims reserving state space models and kalman filter. *The Journal of the Institute of Actuaries*, 110:157Ű181.

Denison, D., Holmes, C., Mallick, B., and Smith, A. (2002). *Bayesian methods for nonlinear classi cation and regression*. John Wiley and Sons.

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2005). *Actuarial theory for dependent risks*. Wiley Online Library.

Dong, X. and Chan, J. (2013). Bayesian analysis of loss reserving using dynamic models with generalized beta distribution. *Insurance: Mathematics and Economics*, 53:355–365.

Embrechts, P. (2009). Copulas: A personal view. *Journal of Risk and Insurance*, 76(3):639–650.

Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modeling extreme events for insurance and nance*. Springer, Berlin.

England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8:443Ů518.

Engle, R. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics,*, 22:367–381.

Fisher, W. and Lange, J. (1973). Loss reserve testing: a report year approach. *Proceedings of the Casualty Actuarial Society*, 60(1):189–207.

Frees, E. and Wang, P. (2006). Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics*, 38(2):360–373.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Genest, C. and MacKay, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *American Statistician,*, 40:280–283.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1996). *Practical Markov Chain Monte Carlo*. Chapman-Hall, New York.

Gilks, W. R., Roberts, G. O., and Sahu, S. K. (1998). Adaptive markov chain monte carlo through regeneration. *Journal of the American Statistical Association*, 93:1045Ű1054.

Girolami, M. and Calderhead, B. (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214.

Gisler, A. (2006). The estimation error in the chain-ladder reserving method: a bayesian approach. *Astin Bulletin*, 36:554–565.

Gogol, D. (1993). Using expected loss ratios in reserving. *Insurance: Mathematics and Economics*, 12(3):297–299.

Goovaerts, M., Dhaene, J., and De Schepper, A. (2000). Stochastic upper bounds for present value functions. *Journal of Risk and Insurance Theory*, 67:1–14.

Guermat, C. and Harris, R. (2002). Forecasting value at risk allowing for time variation in the variance and kurtosis of portfolio returns. *International Journal of Forecasting*, 18:409–419.

Guermat, C. and Harris, R. D. F. (2001). Robust conditional variance estimation and value-at-risk. *Journal of Risk*, 4:25–41.

Gupta, A. and Nagar, D. (2000). *Matrix variate distributions*, volume 104. Chapman & Hall/CRC.

Gyorgy, S. and Shaw, W. (2008). Quantile mechanics. *European journal of applied mathematics*, 19:87–112.

H Aastrup, S. and Arjas, E. (1996). Claims reserving in continuous time; a non-parametric bayesian approach. *ASTIN Bulletin*, 26:139–164.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, 7:223–242.

Haario, H., Saksman, E., and Tamminen, J. (2005a). Componentwise adaptation for high dimensional mcmc. *Computational Statistics*, 20(2):265–273.

Haario, H., Saksman, E., and Tamminen, J. (2005b). Componentwise adaptation for high dimensional mcmc. *Computational Statistics*, 20:265Ű273.

Haastrup, S. and Arjas, E. (1996). Claims reserving in continuous time: A nonparametric bayesian approach. *ASTIN Bulletin*, 26:139–164.

Haberman, S. and Renshaw, A. (1996). Generalized linear models and actuarial science. *The Statistician*, 45:407–436.

Hamilton, J. (1994). *Handbook of Econometrics*. Princeton University Pres.

Happ, S. and Wuthrich, M. (2011). Paid-incurred chain reserving method with dependence modeling. *Astin Bulletin*, To appear.

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109.

Hertig, J. (1985). A statistical approach to the ibnr-reserves in marine insurance. *Astin Bulletin*, 15(2):171–183.

Hossack, I., Pollard, J., and Zenwirth, B. (1999). *Introductory Statistics with Applications in General Insurance*. U. Press, Cambridge.

Hu, W. and Kercheval, A. N. (2008). The skewed t distribution for portfolio credit risk. In *Econometrics and risk management*, volume 22 of *Adv. Econom.*, pages 55–83. Emerald/JAI, Bingley.

Hu, Y., Grimacy, R., and Lian, H. (2012). Bayesian quantile regression for single-index models. *Statistics and Computing*, 23:437–454.

IAA (2009). Measurement of liabilities for insurance contracts: Current estimates and risk margins. *An International Actuarial Research Paper prepared by the ad hoc Risk Margin Working Group, International Actuarial Association*.

Jong, P. D. (2012). Modeling dependence between loss triangles. *North American Actuarial Journal*, 16:74–86.

Kaas, R., Dhaene, J., and M., G. (2000). Upper and lower bounds for sums of random variables. *Mathematics and Economics,*, 27:151–168.

Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, 90:773Ű795.

Koenker, R. and Basse, G. (1978). Regression quantiles. *Econometrica*, 46:33–50.

Koenker, R. and Machado, A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94:1296–1310.

Koop, G. and Potter, S. (1999). Dynamic asymmetries in us unemployment. *Journal of Business and Economic Statistics*, 17:298Ű313.

Kurowicka, D. and Joe, H. (2010). *Dependence Modeling: Vine Copula Handbook*. World Scientific Publishing Co. Pte. Ltd.

Leonard, T. and Hsu, J. (1992). Bayesian inference for a covariance matrix. *The Annals of Statistics*, 20(4):1669–1696.

Li, W. and Lam, K. (1995). Modelling the asymmetry in stock returns by a threshold arch model. *Journal of the Royal Statistical Society: Series B*, 44:333Ű341.

Ling, S. (1999). On the probabilistic properties of a double threshold arma conditional heteroskedasticity model. *Journal of Applied Probability*, 36:688–705.

Liu, J. and Wu, Y. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, pages 1264–1274.

Mack, T. (1991). A simple parametric model for rating automobile insurance or estimating ibnr claims reserves. *Astin Bulletin*, 21:93Ű109.

Mack, T. (1993a). Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin*, 23(1):213–225.

Mack, T. (1993b). Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin*, 23(2):213–225.

Marshall, K., Collings, S., Hodson, M., and O'Dowd, C. (2008). A framework for assessing risk margins. *Prepared by the Risk Margins Task Force for Institute of Actuaries of Australia, 16-th General Insurance Seminar, 9-12 November, 2008, Coolum, Australia.*

Martinez, M. D., Nielsen, J. P., and Verrall, R. J. (2012). Double chain-ladder. *ASTIN Bulletin*, 42:59–76.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall, New York.

McDonald, J. (1984). Some generalized functions for the size distribution of income. *Econometrica*, 52:647–663.

McDonald, J. and Butler, R. (1987). Some generalized mixture distributions with an application to unemployment duration. *The Review of Economics and Statistics*, 69:232Ű240.

McDonald, J. and Newey, W. (1988). Partially adaptive estimation of regression models via the generalized t distribution. *Econometric Theory*, 4:428–457.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management*. Princeton Series in Finance. Princeton University Press, Princeton, NJ. Concepts, techniques and tools.

Meng, X. and Van Dyk, D. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika*, 86(2):301–320.

Merz, M. and Wüthrich, M. V. (2007). Prediction error of the chain ladder reserving method applied to correlated run-off triangles. *Annals of Actuarial Science*, 2:25Ů50.

Merz, M. and Wüthrich, M. (2010). Estimation of tail factors in the paid-incurred chain reserving method. *Submitted preprint*.

Merz, M. and Wuthrich., M. (2010). Paid-incurred chain claims reserving method. *Insurance: Mathematics and Economics*, 46(3):568–579.

Merz, M. and Wüthrich, M. (2010). Paid-incurred chain claims reserving method. *Insurance: Mathematics and Economics*, 46(3):568–579.

Metropolis, N., Rosenbluth, A., and Rosenbluth, M. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1091.

Min, A. and Czado, C. (2010). Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, 8(4):511–546.

Montgomery, A., Zarnowitz, V., Tsay, R., and Tiao, G. (1998). Forecasting the us unemployment rate. *Journal of the American Statistical Association*, 93:478Ű493.

Nelder, J. and Wedderbu, R. (1972). Generalized linear mode. *.ŠŠ Journal of the Royal Statistical Society*, A A135:370–84.

Nelsen, R. (2006a). *An introduction to copulas.* Springer Verlag.

Nelsen, R. (2006b). *An introduction to copulas*. Springer Verlag.

Nelson, C. and Siegel, A. (1987). Parsimonious modeling of yield curves. *Journal of Business*, 60:473–489.

Nelson, R. B. (2006). *An Introduction to Copulas*. Springer.

Neuhaus, W. (2004). On the estimation of outstanding claims. In *Conference paper, presented at the 35th International ASTIN Colloquium*.

Ntzoufras, I. and Dellaportas, P. (2002). Bayesian modeling of outstanding liabilities incorporating claim count uncertainty. *North American Actuarial Journal*, 6:113Ű128.

Patton, A. (2009). Copula–based models for financial time series. *Handbook of financial time series*, pages 767–785.

Paulson, A. and Faris, N. (1985). *A practical approach to measuring the distribution of total annual claims*. Kluwer Academic Publishers, Norwell, MA.

Peters, G., Briers, M., Shevchenko, P., and Doucet, A. (2011a). Calibration and filtering for multi factor commodity models with seasonality: incorporating panel data from futures contracts. *Methodology and Computing in Applied Probability*, pages 1–34.

Peters, G., Byrnes, A., and Shevchenko, P. (2011b). Impact of insurance for operational risk: Is it worthwhile to insure or be insured for severe losses? *Insurance: Mathematics and Economics*, 48:287–303.

Peters, G., Shevchenko, P., and Wuthrich, M. (2009). Model uncertainty in claims reserving within tweedie compound poisson models. *ASTIN Bulletin*, 39:1–33.

Peters, G., Shevchenko, P., Young, M., and Yip, W. (2011c). Analytic loss distributional approach models for operational risk from the stable doubly stochastic compound processes and implications for capital allocation. *Insurance: Mathematics and Economics*, 49:565–579.

Peters, G., Targino, R., and Shevchenko, P. (2013). Understanding operational risk capital approximations: First and second orders. *Governance and Regulation (Invited Special Issue 8th International conference International Competition in Banking: Theory and Practice", Sumy, Ukraine)*, 2:58–79.

Peters, G., Wüthrich, M., and Shevchenko, P. (2010). Chain ladder method: Bayesian bootstrap versus classical bootstrap. *Insurance: Mathematics and Economics*, 47(1):36–51.

Peters, G. W., K. B. L. B. and Mellen, C. (2010). Model selection and adaptive markov chain monte carlo for bayesian cointegrated var models. *Bayesian Analysis*, 5:465–492.

Peters, G. W., K. B. L. B. M. C. and Godsill, S. (2011). Bayesian cointegrated vector autoregression models incorporating alpha-stable noise for inter-day price movements via approximate bayesian computation. *Bayesian Analysis*, 6:755–792.

Posthuma, B., Cator, E., Veerkamp, W., and van Zwet, E. (2008). Combined analysis of paid and incurred losses. In *Casualty Actuarial Society E-Forum, Fall 2008*, page 272.

Quarg, G. and Mack, T. (2004). Munich chain ladder. *mathematik*, 26(4):597–630.

Ramlau-Hansen, H. (1988). A solvency study in non-life insurance. part 1. analysis of fire, windstorm, and glass claims. *Scandinavian Actuarial Journal*, page 3Ű34.

Renshaw, A. and Verrall, R. (1998). A stochastic model underlying the chain ladder technique. *British Actuarial Journal*, 4:903–923.

Roberts, G. and Rosenthal, J. (2009). Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18(2):349–367.

Roberts, G. and Rosenthal., J. (2009). Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18:349Ű367.

Rosenthal, J. et al. (2008). Optimal proposal distributions and adaptive mcmc.

Sawkins, R. (1979a). Analysis of claim run off data–a broad view. *Institute of Actuaries of Austraha General Insurance Seminar l*, pages 30–60.

Sawkins, R. (1979b). Methods of analysing claim payments in general insurance. *Transactions of the Institute of Actuaries of Australia*, (3):435–519.

Scollnik, D. P. M. (2002). Implementation of four models for outstanding liabilities in winbugs: A discusion of a paper by ntzoufras and dellaporta. *North American Actuarial Journal*, 6:113–128.

Sklar, A. (1973). Random variables, joint distribution functions and copulas. *Kybernetika*, 9:449Ű460.

Smith, A. and Roberts, G. (1993). Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B*, 55:3Ű23.

Spiegelhalter, D., Best, N., Carlin, B., and Van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, 64:583–616.

Stacy, E. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, 33:1187–1192.

Stoica, P. and Moses, R. (1997). *Introduction to spectral analysis*, volume 89. Prentice Hall Upper Saddle River, NJ.

Tang, A. and Valdez, E. (2005). Economic capital and the aggregation of risks using copulas. *School of Actuarial Studies and University of New South Wales*, Working Paper:1120–1134.

Tanner, M. and Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, pages 528–540.

Taylor, G. (1977). Separation of inflation and other eliects from the distribution of non-life insurance claim delays. *Astin Bulletin*, IX:219–230.

Taylor, G. (2000). *Claim Reserving. An Actuarial Perspective*. Elsevier Science Publishers, New York.

Taylor, G. (2006). Apra general insurance risk margins. *APRA*.

Taylor, G. and McGuire, G. (2004). Loss reserving with glms: a case study. in: Casualty. *Actuarial Society 2004 Discussion Paper Program*, page 327Ű392.

Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A., and Tuerlinckx, F. (2011). Visualizing distributions of covariance matrices.

Tong, H. (1978). *On a threshold models. In: Chen, C.H. (Ed.), Pattern Recognition and Signal Processing*. Sijthoff and Noordhoff, Amsterdam.

Van Dyk, D. and Meng, X. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1):1–50.

Verbeek. H, G. (1972). An approach to the analysis of claims expenence in motor liability excess of loss reinsurance. *ASTIN Bulletin*, VI:195–202.

Verrall, R. (1989). A state space representation of the chain ladder linear model. *Journal of the Institute of Actuaries*, 116:589Ű609.

Verrall, R. (2000). An investigation into stochastic claims reserving models and the chain- ladder technique. *Insurance: Mathematics and Economics*, 26:91–99.

Verrall, R. (2004). A bayesian generalized linear model for the bornhuetter-ferguson method of claims reserving. *North American Actuarial Journal , 8 North American Actuarial Journal*, 8:67–89.

Verrall, R. and England, P. (2005). Incorporating expert opinion into a stochastic model for the chain-ladder techniqu. *Insurance: Mathematics and Economics*, 37:355–370.

Verrall, R. and Wuthrich, M. (2014). Reversible jump markov chain monte carlo method for parameter reduction in claims reserving. *North American Actuarial Journal*, To appear.

Waszink, H. (2013). Considerations on the discount rate in the cost of capital method for the risk margin. *working paper*.

Wĺuthrich, V., M., and Merz, M. (2008). *Stochastic Claims Reserving Methods in Insurance*. Wiley.

Yang, X., Frees, E., and Zhang, Z. (2011). A generalized beta copula with applications in modeling multivariate long-tailed data. *Insurance: Mathematics and Economics*, 49:265Ű284.

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

Yu, K. and Zhang, J. (2005). A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics theory and Methods*, 34:1867–1879.

Zehnwirth, B. (1994). Probabilistic development factor models with applications to loss reserve variability, prediction intervals, and risk-based capital. *Casualty Actuarial Society Forum, 1994 Spring Forum*, pages 447–605.

Zhang, Y. and Dukic, V. (2014). Predicting multivariate insurance loss payments under the bayesian copula framework. *Journal of Risk and Insurance (forthcoming)*.

Zhang, Y., Dukic, V., and Guszcza, J. (2012). A bayesian nonlinear model for forecasting insurance loss payments. *Journal of the Royal Statistical Society A*, 175:1–20.