Insurance: Mathematics and Economics 59 (2014) 258–278

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

# A copula based Bayesian approach for paid–incurred claims models for non-life insurance reserving

Gareth W. Peters [a,b,*], Alice X.D. Dong [c], Robert Kohn [d]

[a] Department of Statistical Science, University College London UCL, London, UK
[b] CSIRO Mathematics, Informatics and Statistics, Sydney, Australia
[c] University of Sydney, School of Mathematics and Statistics, Australia
[d] UNSW Australian School of Business, Australia

ARTICLE INFO

*Article history:*
Received December 2012
Received in revised form
May 2014
Accepted 26 September 2014
Available online 16 October 2014

*Keywords:*
Chain ladder
Claims reserving
Adaptive Markov chain Monte Carlo

ABSTRACT

Our article considers 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, combining the claims reserving models of Hertig (1985) and Gogol (1993). In the process we extend the independent log-normal model of Merz and Wüthrich (2010) by incorporating different dependence structures using a Data-Augmented mixture Copula paid–incurred claims model.

In this way the paper makes two main contributions: firstly we develop an extended class of model structures for the paid–incurred chain ladder models where we develop precisely the Bayesian formulation of such models; secondly we explain how to develop advanced Markov chain Monte Carlo sampling algorithms to make inference under these copula dependence PIC models accurately and efficiently, making such models accessible to practitioners to explore their suitability in practice. In this regard the focus of the paper should be considered in two parts, firstly development of Bayesian PIC models for general dependence structures with specialised properties relating to conjugacy and consistency of tail dependence across the development years and accident years and between Payment and incurred loss data are developed. The second main contribution is the development of techniques that allow general audiences to efficiently work with such Bayesian models to make inference. The focus of the paper is not so much to illustrate that the PIC paper is a good class of models for a particular data set, the suitability of such PIC type models is discussed in Merz and Wüthrich (2010) and Happ and Wüthrich (2013). Instead we develop generalised model classes for the PIC family of Bayesian models and in addition provide advanced Monte Carlo methods for inference that practitioners may utilise with confidence in their efficiency and validity.



## 1. Introduction

As discussed in Merz and Wüthrich (2010) the main task of reserving actuaries is to predict ultimate loss ratios and outstanding loss liabilities. That is, 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. 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 can be difficult to determine. Therefore, it is important to consider the development of other unified prediction frameworks for the outstanding loss liabilities. Early attempts at such unified combining methods go back to the Munich chain ladder method introduced by Quarg and Mack (2004) which 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.

* Corresponding author at: Department of Statistical Science, University College London UCL, London, UK.
*E-mail address:* gareth.peters@ucl.ac.uk (G.W. Peters).

http://dx.doi.org/10.1016/j.insmatheco.2014.09.011

As a consequence recently there have been new attempts to develop models to combine these two important sources of information, the payment and incurred loss data, into a consistent claims reserving model. The approach proposed in Merz and Wüthrich (2010) is known as a sub-family of the paid–incurred chain ladder (PIC) class of models. As a first instance of such a PIC model, Merz and Wüthrich (2010) introduced a log-normal PIC model and used Bayesian methods to estimate the missing (future) part of the claims reserving triangles based on both payment and loss incurred information. The major advantage of the PIC model structure is that the full predictive distribution of the outstanding loss liabilities can be quantified. One important limitation of the model of Merz and Wüthrich (2010) is that it does not develop the dependence properties of the PIC model that may be applicable or of interest to consider in loss reserving data observed in practice.

The first attempt to address the incorporation of dependence was recently proposed in Happ and Wüthrich (2013) where a restricted class of models was proposed for a very simple dependence structure, see Happ and Wüthrich (2013, Figure 1.1). This model is parsimonious but was limited to only three parameters for the correlations which were not incorporated into a formal Bayesian estimation approach, and instead fixed deterministically *a priori* via some hybrid of Bayesian and frequentist methods.

Our article extends the proposed Bayesian PIC models to capture a more flexible range of dependence structures that includes as special cases the model classes of Merz and Wüthrich (2010) and Happ and Wüthrich (2013). We note that in our framework we also show how to extend the model of Happ and Wüthrich (2013) into a complete Bayesian estimation framework rather than a hybrid frequentist and Bayesian approach that they considered. We aim to significantly enhance the class of dependence model structures one may consider in the PIC setting whilst sticking to a complete and formal Bayesian formulation of the problem. Achieving this is non-trivial in both the development of the Bayesian model structures and the estimation under such models. We note that in some cases we will show that this can also be done in a parsimonious manner under mixture Archimedean copula structures. This will be particularly relevant if only vague *a-priori* information is known about the PIC model parameters and the loss payment and incurred loss triangles are small.

In general one may consider three forms of dependence, namely, dependence within payment data, within incurred loss data, and even between payment and incurred loss data. In general it will be up to practitioners as to which of these forms of dependence they find most practically relevant in practice. The intention of this paper is not to advocate that dependence is always present, instead we provide a general class of PIC Bayesian models that can accommodate a wide array of dependence structures as well as a suitable estimation framework based on MCMC methods that practitioners can utilise to explore aspects of their given data accurately.

It is important to recognise that in this paper we are not attempting to argue for or against particular model structures or dependence features in the PIC family of models for application in practice. Instead we simply demonstrate carefully how to develop generalised classes of such models that are complete in a Bayesian sense and preserve certain important statistical assumptions such as consistency of the tail dependence assumptions throughout the development and accident year structures in the likelihood, even in the presence of unobserved components as arises in the triangle structures of the payment and incurred loss triangles.

## 2. Contributions: extending dependence structures for PIC Bayesian models

The main focus of this paper is to develop new flexible classes of copula dependent PIC Bayesian models with appropriate non-trivial inference procedures that will allow practitioners and actuaries to explore in their given applications this class of models in order to test their worth in practical settings. To be able to utilise efficiently general dependence structures in PIC Bayesian models in practice we introduce to the actuarial literature the data augmentation method which is an auxiliary variable framework that is not previously utilised in the actuarial literature. Given these extended Bayesian PIC models we also provide the appropriate MCMC methods to handle these new PIC Bayesian models, since standard MCMC methods will not be adequate to efficiently work with such models in practice. This is an important contribution to introduce to the actuarial literature modern classes of adaptive MCMC and auxiliary variable methods recently developed in statistics that actuaries can then explore for their given applications. In particular we make two main contributions to the literature for this class of PIC models.

**Contribution 1:** We extend the class of dependence structures available for the PIC reserving models in the process making the inference procedure into a consistent and coherent fully Bayesian formulation. This is unlike previous proposed approaches in Happ and Wüthrich (2013) which have only very simple and restrictive correlation based dependence structures which were highly constrained in the form of dependence between the development and accident years of the payments and incurred claims triangles and which did not allow for the possibility of interesting features such as tail dependence. In addition we note that we achieve this also in more sophisticated dependence structures in some cases without incurring an increase in the number of parameters relative to the simplified model of Happ and Wüthrich (2013). In other cases since our approach produces a complete and consistent fully Bayesian formulation we may utilise prior beliefs to inform the *a-priori* belief in a particular form of dependence which can therefore still be estimated sensibly in applications where the size of the loss reserving triangle is not large. Then since we are developing complete Bayesian models, practitioners can utilise standard model selection methods to select appropriate dependence structures for their data and application.

There are two technical difficulties we address when extending the dependence structures utilised in the PIC reserving models within a formal complete Bayesian framework. The first occurs when working with general linear dependence structures in the PIC model likelihood, as encoded by covariance-correlation structures between rows (columns) of the payment or incurred or both loss data triangles. The challenge involves being able to specify and evaluate in closed form "point-wise" the PIC Bayesian posterior model (up to proportionality). This is challenging as we need to ensure the prior and posterior admit appropriate restrictions on the support upon which they are defined to produce densities for different structures of positive definite symmetric covariance matrices (with constraints). The second problem occurs when other forms of dependence are considered such as those obtained when considering copula dependence models not purely obtained from a covariance-correlation matrix structure.

We solve the first challenge through development of a generic PIC Bayesian model with special specification of a class of matrix-variate Inverse-Wishart priors defined over the space of positive definite matrices that will admit conditional posterior conjugacy structures under our family of PIC Bayesian models. The second problem is addressed by developing models for general copula dependence structures in the likelihood that will produce closed form posterior densities even in the presence of only partial observation information in each column or row of the payment and incurred claims data sets. This problem is specified in a class of PIC Bayesian models with the joint likelihood over the payment and incurred loss data triangles which has a mixture copula structure defined over the observed payments and incurred losses in each accident year row (or column) of the reserving matrix. This last point is non-trivial to achieve if one wishes to preserve the same tail dependence features across the predictions to be consistent with the observed payments or incurred losses that have developed so far to that point. This copula "consistency" is defined with respect to for instance an entire row (or column) of for instance the payment data likelihood when the likelihood for that row is only marginally observed in a given set of columns. This can be stated as the challenge of specifying a family of Bayesian models for PIC likelihood structures which will ensure that the likelihood evaluated on a partially observed sub-vector of the observations will still produce a consistent dependence structure for the entire row, i.e. predictions, whilst also admitting a consistent and tractable Bayesian posterior distribution conditioned only upon the marginal information observed.

The incorporation of 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 for what we consider an illustrative toy model:

(i) an independent payment data model;
(ii) the independent payment and incurred claims data model of Merz and Wüthrich (2010);
(iii) a novel dependent lag-year telescoping block diagonal Gaussian copula payment and incurred claim data model incorporating conjugacy via transformation. Note: 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;
(iv) a novel data-augmented mixture Archimedean copula dependent payment and incurred claim data model. This involves a mixture of Clayton and Gumbel copulas for upper and lower tail dependence features in the development years for payments and incurred losses.

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 (2013).

Such extended models will allow practitioners to further explore questions such as when it may be suitable to consider a fixed covariance structure to describe the correlation between payment and incurred loss which would assume that the correlations between different development periods are identical. By extending the class of PIC Bayesian models we allow practitioners to also consider that perhaps correlations differ across development periods for various reasons, such as different stages of the life cycle for a claim and internal policy changes. So for instance in our extended class of PIC Bayesian models, in order to fully incorporate such correlation structures, we introduce a block covariance structure to allow for the variation between different development periods within payment and incurred losses. In addition, we also develop a class of PIC Bayesian models that allows for the possibility for practitioners to explore upper and lower asymmetric tail dependence features within the PIC Bayesian model class so that practitioners may also consider testing for such features in their applications.

We note that in practice using our methods actuaries can now make accurate complete Bayesian inference based on their *a-priori* beliefs in model structure and dependence structure and test for such model features in their given applications via standard Bayesian model selection methods such as Bayes factor ranking or information criteria such as BIC or DIC. As such this is not at all the focus of this paper, we simply aim to provide the tools for actuaries to have confidence to get to the point of being able to undertake this model analysis accurately and efficiently in a proper full Bayesian formulation.

**Contribution 2:** We develop efficient and accurate Markov chain Monte Carlo (MCMC) sampling methods for practitioners to work with these classes of PIC Bayesian models. This is challenging as it involves designing Markov transition kernels used in the MCMC to sample from the posterior, which involve generation of positive definite matrices from the posterior under different constraint structures, such as block-wise telescoping covariance structures. In particular we develop a family of adaptive Markov chain samplers that restricts the proposed Markov chain states to remain on the manifold of such matrices.

In the case of general copula dependence structures such as mixtures of Archimedean copulas we preserve the dependence structure in the joint data likelihood as well as in the posterior predictive for the reserves 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.

## 3. Review of the Merz–Wüthrich independence copula paid–incurred claims model

This section introduces the first independent 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 Wüthrich (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 summarised in triangular form which is utilised for both the claims payments and the incurred losses, as presented in Fig. 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) $P_{0:J,0:j} = \left\{P_{k,l} : 0 \le k \le J, 0 \le l \le j\right\}$. (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 $\mathrm{diag}(A, B, C) = \mathrm{diag}(\mathrm{diag}(A, B), C)$ 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 $\delta_i$.

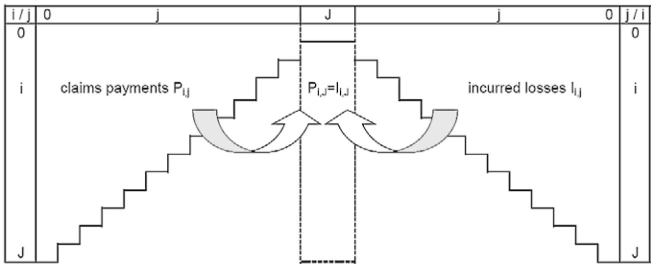


**Fig. 1.** Claims triangle for payment data and incurred data, including constraint on zero case estimates at development period $J$.
*Source:* Merz and Wüthrich (2010).

(vi) Define the vectorisation operator on a $p \times n$ matrix $A$, denoted by Vec($A$), as the stacking of the columns to create a vector.

As in Merz and Wüthrich (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 3.1** (*Independent PIC Log-Normal (Model I)*). *The model assumptions for the independent model of Merz and Wüthrich (2010) are:*

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

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

- *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 $(\xi_{0,0}, \dots, \xi_{J,J}, \zeta_{0,0}, \dots, \zeta_{J,J-1})$ has independent components with*

$$\xi_{i,j} \sim N\left(\Phi_j,\ \sigma_j^2\right) \quad \text{for } i \in \{0, \dots, J\} \text{ and } j \in \{0, \dots, J\},$$
$$\zeta_{k,l} \sim N\left(\Psi_l,\ \tau_l^2\right) \quad \text{for } k \in \{0, \dots, J\} \text{ and } l \in \{0, \dots, J-1\}.$$

- *The parameter vector for the model is $\mathbf{\Theta} = \left(\Phi_0, \dots, \Phi_J, \Psi_0, \dots, \Psi_{J-1}, \sigma_0, \dots, \sigma_J, \tau_0, \dots, \tau_{J-1}\right)$. It is assumed that the components of $\mathbf{\Theta}$ are independent a priori. The prior density for $\mathbf{\Theta}$ has independent components, with $\sigma_j$, $\tau_j$ both positive for all j.*
- *It follows that*

$$\log\left(\frac{P_{i,j}}{P_{i,j-1}}\right) \sim N\left(\Phi_j,\ \sigma_j^2\right) \quad \text{and} \\ \log\left(\frac{I_{i,j}}{I_{i,j+1}}\right) \sim N\left(-\Psi_l,\ \tau_l^2\right). \tag{3.1}$$

Let $\{\mathbf{P}, \mathbf{I}\} = \{P_{i,j}, I_{k,l};\ 0 \le i, j, k, \le J,\ 0 \le l \le J-1\}$. Then, based on Model Assumptions 3.1 and the observed matrices $P$ and $I$, the likelihood for $\mathbf{\Theta}$ is given by three components, see derivation in Merz and Wüthrich (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 Eq. (3.2) given in Box I.

As noted in Merz and Wüthrich (2010), there are several consequences of the model assumptions made regarding the restriction $I_{i,J} = P_{i,J}$ which applies for all $i \in \{1, 2 \dots, J\}$. 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 (2013).

Merz and Wüthrich (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 Wüthrich (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 3.2 summarises their results.

**Lemma 3.2.** *The relationships between consecutive payment development year losses in a given accident year is given conditionally according to*

$$\left[\log\left(\frac{P_{i,j}}{P_{i,j-1}}\right) \middle| P_{0:J,0:j-1}, \mathbf{\Theta}\right] \sim \mathcal{N}\left(\Phi_j, \sigma_j^2\right), \quad \forall j \ge 0 \tag{3.3}$$

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

$$\mathbb{E}\left[P_{i,j} | P_{0:J,0:j-1}, \mathbf{\Theta}\right] = P_{i,j-1} \exp\left(\Phi_j + \sigma_j^2/2\right). \tag{3.4}$$

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

$$\left[\log\left(I_{i,j+l}\right) \middle| I_{0:J,0:j-1}, I_{i,J}, \mathbf{\Theta}\right] \\ \sim \mathcal{N}\left(\mu_{j+1} + \frac{\nu_{j+1}^2}{\nu_j^2}\left(\log(I_{i,j}) - \mu_j\right), \nu_{j+1}^2(1 - \nu_{j+1}^2/\nu_j^2)\right). \tag{3.5}$$

*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 Wüthrich (2010).*

Furthermore, for all accident years $i \in \{1, 2, \dots, J\}$, 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 Eq. (3.6), which are given by Merz and Wüthrich (2010, Equation 1.1) as,

$$E\left[P_{i,J} \middle| \mathbf{\Theta}\right] = E\left[I_{i,J} \middle| \mathbf{\Theta}\right] = \exp\left(\sum_{m=0}^{J} \Phi_m + \sigma_m^2/2\right). \tag{3.6}$$

## 4. Extended Gaussian copula structures for PIC Bayesian models

This section discusses an important aspect of extending the original Log-Normal PIC model of Merz and Wüthrich (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 $i, j, k, l \in \{1, 2, \dots, J\}$ 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 general we note that 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 previously to be important in simple cases

$$f(\mathbf{P},\mathbf{I}|\,\boldsymbol{\Theta}) = \underbrace{\prod_{j=0}^{J}\prod_{i=0}^{J-j}\frac{1}{\sqrt{2\pi}\,\sigma_j P_{i,j}}\ \exp\left\{-\frac{1}{2\sigma_j^2}\left(\Phi_j-\log\left(\frac{P_{i,j}}{P_{i,j-1}}\right)\right)^2\right\}}_{\text{Component1: payment}}$$
$$\times\underbrace{\prod_{i=1}^{J}\frac{1}{\sqrt{2\pi\,(\upsilon_{J-i}^2-\omega_{J-i}^2)}I_{i,J-i}}\ \exp\left\{-\frac{1}{2(\upsilon_{J-i}^2-\omega_{J-i}^2)}\left(\mu_{J-i}-\eta_{J-i}-\log\left(\frac{P_{i,J-i}}{P_{i,J-i}}\right)\right)^2\right\}}_{\text{Component2: Discounted final development year restricted payment and incurred}}$$
$$\times\underbrace{\prod_{j=0}^{J-1}\prod_{i=0}^{J-j-1}\frac{1}{\sqrt{2\pi}\,\tau_j I_{i,j}}\ \exp\left\{-\frac{1}{2\tau_j^2}\left(-\Psi_j+\log\left(\frac{I_{i,j}}{I_{i,j+1}}\right)\right)^2\right\}}_{\text{Component3: incurred}}. \tag{3.2}$$

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 $\mu_j=\sum_{m=0}^{J}\Phi_m-\Sigma_{n=j}^{J-1}\varphi_n$.

**Box I.**

by Happp and Wüthrich (2013) so we use this study as motivation to develop a more complete picture of the PIC Bayesian model classes and inference approaches.

### *4.1. Dependence via payment loss ratios and incurred loss ratios (Bayesian model II)*

We use a complete Bayesian approach, based on results in Lemma A.2 and Model Assumptions 4.1, 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 Lemmas A.2 and A.3.

**Model Assumptions 4.1** (*Dependent Payment–Incurred Ratios: PIC Log-Normal (Model II)*)**.** *The model assumptions for the Bayesian 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 ith development year, denoted by $\Xi_i=\left(\xi_{i,0},\xi_{i,1},\zeta_{i,1},\xi_{i,2},\zeta_{i,2},\ldots,\xi_{i,J},\zeta_{i,J}\right)$, is assumed distributed according to an inverse Wishart distribution prior (see definition and properties in Lemmas A.2 and A.3),*
$$\Sigma_i\sim\mathcal{IW}(\Lambda_i,k_i) \tag{4.1}$$
*where $\Lambda_i$ is a $((2J+1)\times(2J+1))$ positive definite matrix and $k_i>2J$.*
- *Conditionally, given $\boldsymbol{\Theta}=\left(\Phi_0,\ldots,\Phi_J,\Psi_0,\ldots,\Psi_J\right)$ and the $(2J+1)\times(2J+1)$-dimensional covariance matrix $\Sigma$, 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 $\Xi$ 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 $f_{\Xi}^{MVN}(\Xi|M,\Sigma,\Omega)$, see the definition and properties in Lemma A.1. The sufficient matrices are then the $((2J+1)\times(J+1))$ mean matrix $M=\left[\boldsymbol{\Theta}',\ldots,\boldsymbol{\Theta}'\right]$, column dependence given by $((2J+1)\times(2J+1))$ dimensional covariance matrix $\Sigma$ and row dependence given by $((J+1)\times(J+1))$ dimensional matrix $\Omega$. If $\Omega=\mathbb{I}_{J+1}$, the covariance of the vectorisation of $\widetilde{\Xi}=\text{Vec}(\Xi)$ is*
  $$\widetilde{\Sigma}=\mathbb{C}\text{ov}\left(\widetilde{\Xi}\right)=\bigoplus_{i=0}^{J}\Sigma_i=\begin{pmatrix}\Sigma_0&\mathbf{0}&\cdots&\mathbf{0}\\ \mathbf{0}&\Sigma_1&\cdots&\mathbf{0}\\ \vdots&\ddots&\ddots&\vdots\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\Sigma_J\end{pmatrix}, \tag{4.2}$$
  *where it is assumed in the model in Happ and Wüthrich (2013) that $\Sigma_i=\mathbb{C}\text{ov}(\Xi_i)=\Sigma$. However, this need not be the case and it is possible to consider two extensions, the first in which $\mathbb{C}\text{ov}(\Xi_i)$ varied as a function of $i\in\{0,1,\ldots,J\}$ and the second being the most general of these model structures given by the assumption*
  $$\mathbb{C}\text{ov}\left(\widetilde{\Xi}\right)=\Sigma\otimes\Omega. \tag{4.3}$$
  - *For all accident years, $i\in\{0,1,\ldots,J\}$, the ultimate payment losses and incurred losses are equal a.s., $P_{i,J}=I_{i,J}$.*
- *The matrix $\widetilde{\Sigma}$ is positive definite and the components of $\boldsymbol{\Theta}$ are independent with prior distributions*
$$\Phi_i\sim\mathcal{N}\left(\phi_i,s_i^2\right)\quad\text{and}\quad\Psi_j\sim\mathcal{N}\left(\psi_j,t_j^2\right), \tag{4.4}$$
*and hyper-prior distributions*
$$s_i^2\sim\mathcal{IG}(\alpha_i,\beta_i)\quad\text{and}\quad t_j^2\sim\mathcal{IG}\left(a_j,b_j\right), \tag{4.5}$$
*for all $i\in\{1,\ldots,J\}$ and $j\in\{0,\ldots,J\}$.*

This model extends the model developed in Happ and Wüthrich (2013) which assumes that $\Sigma$ is fixed and known with a tri-diagonal structure. The extension we consider generalises 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 Wüthrich (2013) where they assume $\Sigma=\Sigma_i,\forall i\in\{0,1,\ldots,J\}$, we also allow for variation in $\Sigma_i$ 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 Lemmas 4.2 and 4.5.

**Lemma 4.2.** *Conditional upon $\Lambda_i$ and $k_i$, for all $i$ in $\{0,1,\ldots,J\}$, and given the marginal distributions for $\Sigma_i$ follow $\Sigma_i\sim\mathcal{IW}(\Lambda_i,k_i)$ with $\Lambda_i$ a $((2J+1)\times(2J+1))$ positive definite matrix and $k_i>2J$, the joint distribution for the $\left((2J^2+3J+1)\times(2J^2+3J+1)\right)$ covariance matrix $\widetilde{\Sigma}$ for the vectorised matrix for $\Xi$, given by $\widetilde{\Xi}=\text{Vec}(\Xi)$, under the assumption of independence between development years,*

$$\widetilde{\Sigma}=\mathbb{C}\text{ov}\left(\widetilde{\Xi}\right)=\bigoplus_{i=0}^{J}\Sigma_i=(\Sigma_0\oplus\cdots\oplus\Sigma_J), \tag{4.6}$$

*results in a joint distribution given by:*

$$\widetilde{\Sigma} \sim \mathcal{IW}\left(\widetilde{\Lambda}, \widetilde{k}\right), \tag{4.7}$$

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

$$\widetilde{\Lambda} = \bigoplus_{i=0}^{J} \Lambda_i. \tag{4.8}$$

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

$$\begin{aligned} \mathbb{E}\left[\widetilde{\Sigma}|\widetilde{\Lambda}, \widetilde{k}\right] &= \frac{1}{\left(\sum_{i=0}^{J} k_i\right) - \left(2J^2 + 3J\right)} \widetilde{\Lambda}, \quad \text{and} \\ m\left(\widetilde{\Sigma}\right) &= \frac{1}{2J^2 + 3J + 1 + \sum_{i=0}^{J} k_i} \widetilde{\Lambda}. \end{aligned} \tag{4.9}$$

*The proof of this result is a consequence of the results in Lemma A.2, the model assumptions and the properties of an inverse Wishart distributions; see Gupta and Nagar (2000) [Chapter 3].* □

**Remark 4.3.** We can demonstrate that under the proposed model assumptions the selection of the factorised covariance structure in Lemma 4.2 produces Bayesian conjugacy in the joint posterior of the model parameters given observed payment losses and incurred losses.

**Remark 4.4.** It is noted in Happ and Wüthrich (2013) and Lemma 4.2 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

$$[\boldsymbol{X}_i|\boldsymbol{\Theta}] = [B_i \Xi_i|\boldsymbol{\Theta}] \sim \mathcal{N}\left(B_i M_i, B_i \Sigma_i B_i'\right), \tag{4.10}$$

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

$$\left[\widetilde{\boldsymbol{X}}|\boldsymbol{\theta}\right] = \left[B\widetilde{\Xi}|\boldsymbol{\Theta}\right] \sim \mathcal{N}\left(B\mathrm{Vec}(M), B\left(\Sigma \otimes \Omega\right)B^T\right). \tag{4.11}$$

Furthermore, if we consider the property of multivariate Gaussian distributions given in Lemma 4.5 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 4.5 below. This is directly relevant for specifying the resulting likelihood model.

**Lemma 4.5.** *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* $\mathbb{C}\mathrm{ov}(\boldsymbol{Y}) = \Sigma$*, and partition* $Y = \left[Y^{(1)'}, Y^{(2)'}\right]'$*. Then the conditional distribution of* $Y^{(1)}$ *given* $Y^{(2)}$ *and the marginal distribution of* $\boldsymbol{Y}^{(1)}$ *is*

$$\left[\boldsymbol{Y}^{(1)}|\boldsymbol{Y}^{(2)}\right] \sim \mathcal{N}\left(\bar{\boldsymbol{\mu}}, \bar{\Sigma}\right), \tag{4.12}$$

*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*

$$\mu = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{2,1} \\ \Sigma_{1,2} & \Sigma_{2,2} \end{pmatrix}. \tag{4.13}$$

Definition 4.6 below defines a family of permutation matrix operators. This permutation family allows the representation of the vectorisation of the two loss triangles under different permutations that facilitate dependence specifications in the proposed models that admit conjugacy.

**Definition 4.6.** Let $Y$ be an $n \times n$ matrix, with $\widetilde{Y} = \left[Y_{1,1}, Y_{1,2}, \ldots\right]'$ and with $\mathrm{Vec}(Y)$ defined as $\mathrm{Vec}(Y) = \left[Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n}, Y_{2,1}, \ldots, Y_{2,n}, \ldots, Y_{n,n}\right]'$. Define the family of permutation matrix operators, denoted by $\mathcal{P}_{\boldsymbol{i}}^*$ and indexed by $p \times 2, p \le n^2$, 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 $\mathcal{P}_{\boldsymbol{i}}^* : \mathrm{Vec}(Y) \mapsto \mathrm{Vec}(Y)^*$ given by

$$\begin{aligned} \mathcal{P}_{\boldsymbol{i}}^*\left(\mathrm{Vec}(Y)\right) &= P_{\boldsymbol{i}}^*\mathrm{Vec}(Y) \\ &= \left[Y_{[\boldsymbol{i}]_1}, Y_{[\boldsymbol{i}]_2}, \ldots, Y_{[\boldsymbol{i}]_p}, \mathrm{Vec}(Y)_{\setminus\boldsymbol{i}}'\right]', \end{aligned} \tag{4.14}$$

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

$$P_{\boldsymbol{i}}^* = P_{\boldsymbol{i}} \oplus \mathbb{I}_{n^2-p} = \begin{bmatrix} P_{\boldsymbol{i}} & 0_{n^2-p,n^2-p} \\ 0_{n^2-p,n^2-p} & \mathbb{I}_{n^2-p} \end{bmatrix}, \tag{4.15}$$

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 4.5, one can state the result in Proposition 4.7 which is based on a generalisation of the result in Happ and Wüthrich (2013, Lemma 2.1) to the model developed above. We consider two cases for the dependence structures in Propositions 4.7 and 4.8.

**Proposition 4.7.** *Consider the ith accident year. Conditional on the model parameters* $\boldsymbol{\Theta}$ *and the covariance matrix of the ith accident year*

$$\Sigma_i = \begin{pmatrix} [\Sigma_i]_{1,1} & [\Sigma_i]_{2,1} \\ [\Sigma_i]_{1,2} & [\Sigma_i]_{2,2} \end{pmatrix}, \tag{4.16}$$

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

$$\left[\boldsymbol{X}_i^{(2)}|\boldsymbol{X}_i^{(1)}, \boldsymbol{\Theta}\right] \sim \mathcal{N}\left(\bar{\boldsymbol{\mu}}^{(2)}, \bar{\Sigma}_i^{(2)}\right) \tag{4.17}$$

*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 4.8** (*Conditional Distribution of Unobserved Payment and Incurred Losses*)**.** *Consider the ith accident year and define indices for this year (vector of tuples), given by matrix* $\boldsymbol{i} = \{(k, j) : \forall j \in \{J - k + 1, \ldots, J\}\} \cup \left\{(k, j) : \forall k \in \{0, 1, \ldots, J\}, j \in \{0, \ldots,\right.$

$J-k\}\big\}$. *Then consider the transformed vector of log payment and incurred losses* $\mathcal{P}_{\boldsymbol{i}}^{*}\left(\widetilde{X}\right)$ *defined by*

$$\mathcal{P}_{\boldsymbol{i}}^{*}\left(\widetilde{X}\right)\sim\mathcal{N}\left(P_{\boldsymbol{i}}^{*}\mathrm{Vec}(M),P_{\boldsymbol{i}}^{*}\left(\Sigma\otimes\Omega\right)(P_{\boldsymbol{i}}^{*})'\right), \tag{4.18}$$

*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 structure* $\widetilde{\Sigma}=\Sigma\otimes\Omega$ *with matrices* $\Sigma$ *and* $\Omega$, *and* $\left[\widetilde{X}^{*}\right]^{(2)}$ *the following results hold:*

- *The conditional distribution for the log of the* unobserved payment losses and incurred losses *in the ith 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*

$$\left[\left[\widetilde{X}^{*}\right]^{(1)}\mid\left[\widetilde{X}^{*}\right]^{(2)},\boldsymbol{\Theta}\right]\sim\mathcal{N}\left(\bar{\boldsymbol{\mu}}^{(1)},\bar{\Sigma}_{i}^{(1)}\right). \tag{4.19}$$

- *The covariance matrix* $\bar{\Sigma}_{i}^{(1)}$ *is the positive 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* $\Gamma$ *and defined by the top subblock of the permuted covariance matrix*

$$P_{\boldsymbol{i}}^{*}\left(\Sigma\otimes\Omega\right)(P_{\boldsymbol{i}}^{*})' =\begin{bmatrix}\Gamma & \left[P_{\boldsymbol{i}}^{*}\left(\Sigma\otimes\Omega\right)(P_{\boldsymbol{i}}^{*})'\right]_{2,1}\\ \left[P_{\boldsymbol{i}}^{*}\left(\Sigma\otimes\Omega\right)(P_{\boldsymbol{i}}^{*})'\right]_{1,2} & \left[P_{\boldsymbol{i}}^{*}\left(\Sigma\otimes\Omega\right)(P_{\boldsymbol{i}}^{*})'\right]_{2,2}\end{bmatrix}. \tag{4.20}$$

- *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* $\Gamma$ *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* $\boldsymbol{\mu}^{(1)}=\left[P_{\boldsymbol{i}}^{*}\mathrm{Vec}(M)\right]_{1:J-i}$ *and the second* $J-i$ *to* $J-i+\left(\sum_{n=-1}^{J}(J-n)\right)$ *elements are given by* $\boldsymbol{\mu}^{(2)}=\left[P_{\boldsymbol{i}}^{*}\mathrm{Vec}(M)\right]_{J-i:J-i+\left(\sum_{n=-1}^{J}(J-n)\right)}$.

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 Eq. (4.21). 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 $\boldsymbol{i}=\{(i,j):\forall i\in\{0,1,\dots,J\},\,j\in\{0,\dots,J-i\}\}$. The resulting likelihood is given by the matrix-variate Gaussian distribution in Eq. (4.21) which is given in Box II.

We note that our proposed models also allow one to consider the dependence structures of Happ and Wüthrich (2013) who assume that $\Sigma_i=\Sigma,\forall i\in\{0,1,\dots,J\}$ and $\Omega=\mathbb{I}_{J+1}$, with the specific setting of $\Sigma$ 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 $\zeta_i^j$ 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 $\rho_0,\rho_1,\rho_2$, respectively. However, the argument for more general dependence structure that were introduced as extensions to the model of Happ and Wüthrich (2013) 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 paper 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.

### *4.2. Dependence between development lag years for payment and incurred losses (Bayesian 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 recognised 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 4.9 below.

**Model Assumptions 4.9** (*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* $\left[\log P_{i,0},\log P_{i,1},\dots,\log P_{i,J-i}\right]$ *in the ith 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 Lemmas A.3 and A.2 ) for each matrix defined according to*

$$\Sigma_i^P\sim\mathcal{IW}\left(\Lambda_i^P,k_i^P\right)\quad\text{and}\quad\Sigma_i^I\sim\mathcal{IW}\left(\Lambda_i^I,k_i^I\right), \tag{4.22}$$

*where* $\Lambda_i^P$ *and* $\Lambda_i^I$ *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:*

$$\widetilde{\Sigma}=\mathbb{C}\mathrm{ov}\left(\left[\log P_{0,0},\dots,\log P_{0,J},\log P_{1,0},\right.\right.$$
$$\log P_{1,J-1},\dots,\log P_{J,0},\log I_{0,0},\dots,$$
$$\left.\left.\log I_{0,J-1},\dots,\log I_{J,0}\right]\right)$$

$$f\left(\left[\widetilde{X}^{*}\right]^{(1)}\middle|\,\mathbf{\Theta},\Sigma,\Omega\right)=\frac{\exp\left[\left(\left[\widetilde{X}^{*}\right]^{(1)}-\left[\mathcal{P}_{\boldsymbol{i}}^{*}\left(\mathrm{Vec}(M)\right)\right]^{(1)}\right)\left[\left[P_{\boldsymbol{i}}^{*}\left(\Sigma\otimes\Omega\right)\left(P_{\boldsymbol{i}}^{*}\right)'\right]^{(1)}\right]^{-1}\left(\left[\widetilde{X}^{*}\right]^{(1)}-\left[\mathcal{P}_{\boldsymbol{i}}^{*}\left(\mathrm{Vec}(M)\right)\right]^{(1)}\right)\right]}{(2\pi)^{\frac{1}{2}\sum_{n=-1}^{J}(J-n)}\left|\left[P_{\boldsymbol{i}}^{*}\left(\Sigma\otimes\Omega\right)\left(P_{\boldsymbol{i}}^{*}\right)'\right]^{(1)}\right|^{\frac{1}{2}\sum_{n=-1}^{J}(J-n)}}. \tag{4.21}$$

**Box II.**

$$= \left(\bigoplus_{i=0}^{J}\Sigma_0^P\right)\oplus\left(\bigoplus_{i=0}^{J}\Sigma_0^I\right)$$

$$\sim \mathcal{IW}\left(\left(\bigoplus_{i=0}^{J}\Lambda_0^P\right)\oplus\left(\bigoplus_{i=0}^{J}\Lambda_0^I\right),\sum_{i=0}^{J}\left(k_i^P+k_i^I\right)\right). \tag{4.23}$$

- *Conditionally, given* $\mathbf{\Theta} = \left(\Phi_0,\ldots,\Phi_J,\Psi_0,\ldots,\Psi_J\right)$ *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 vectorised random matrix of observed payment and incurred losses, with ith column* $\boldsymbol{X}_i \in \mathbb{R}^{2J+1}$ *given by*
    $\boldsymbol{X}_i = \left[\log I_{i,0},\log P_{i,0},\log I_{i,1},\log P_{i,1},\ldots,\log I_{i,J-1},\right.$ $\left.\log P_{i,J-1},\log I_{i,J}\right]$.
    *Then given the matrix of permutation indices* $\boldsymbol{i} = \left[(1,2),(1,4),\ldots,(1,2(J-1)),(2,2),(2,4),\ldots,(2,2J-4),\ldots,(J,1),(1,1),(1,3),\ldots,(J-1,1),(J-1,2)\right]$ *characterising the elements of the marginal distribution for the observations, the transform* $\mathcal{P}_{\boldsymbol{i}}^{*}(\mathrm{Vec}(\boldsymbol{X}))$ *has multivariate Gaussian distribution with covariance structure* $\widetilde{\Sigma}$. *Note,* $\mathcal{P}_{\boldsymbol{i}}^{*}(\mathrm{Vec}(\boldsymbol{X})) = \left[\log P_{0,0},\log P_{0,1},\ldots,\log P_{0,J},\ldots,\log P_{J,0},\right.$ $\left.\log I_{0,0},\ldots,\log I_{0,J-1},\log I_{1,0},\ldots,\log I_{J-1,0},\log I_{J-1,1}\right]$.
  - *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*

$$\Phi_i\sim\mathcal{N}\left(\phi_i,s_i^2\right)\quad\text{and}\quad\Psi_j\sim\mathcal{N}\left(\psi_j,t_j^2\right) \tag{4.24}$$

*and hyper-prior distributions*

$$s_i^2\sim\mathcal{IG}\left(\alpha_i,\beta_i\right)\quad\text{and}\quad t_j^2\sim\mathcal{IG}\left(a_j,b_j\right) \tag{4.25}$$

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

This proposed model is therefore another generalisation of the dependence structure of the model structure proposed in Happ and Wüthrich (2013). As such, the likelihood structure is given by the multivariate Gaussian given in Eq. (4.21) with the covariance matrix given by the telescoping diagonal block size covariance matrix structure in Eq. (4.23).

## 5. Mixture-Archimedean copula structures for PIC Bayesian models: model IV

We next present alternative dependence structures for the PIC Bayesian model that admit tail dependence features in the likelihood for the data. Previous studies on claims reserving that have incorporated copula based models, such as Zhang and Dukic (2013) have done so through regression based frameworks. In this paper we construct a mixture copula allowing for upper and lower tail dependence asymmetry to be considered in the data. To consider such a model structure in the likelihood of the payment and incurred data requires the introduction of auxiliary variables to enable computation and ensure consistency of dependence structures across rows or columns that are partially observed. The approach developed involves modifying the posterior distribution by embedding the target posterior distribution for the model parameters into a higher dimensional support comprised of the original model parameters and 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.

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, canonical 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 used 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 5.3, that is directly comparable to that used for Model II in Assumptions 4.1.

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 B.1 for the characteristics of the Archimedean family of copulas and Lemma B.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 B.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 parameterisations. 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 B.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 (2006, Section 4.3, Table 4.1) and Lemma B.2: the Clayton copula density is denoted by $c^C(u_1,\ldots,u_n;\rho^C)$ with $\rho^C\in[0,\infty)$ the dependence parameter; the Gumbel copula density is denoted by $c^G(u_1,\ldots,u_n;\rho^G)$ with $\rho^G\in[1,\infty)$ the dependence parameter; and the Frank copula density is denoted by $c^F(u_1,\ldots,u_n;\rho^F)$ with $\rho^F\in\mathbb{R}/\{0\}$ the dependence parameter. The upper and lower tail dependence, in terms of copula parameters, for these mixture components are given in Hofert et al. (2012) and the fact that the mixture copula is still a copula is provided in Lemma 5.1.

**Lemma 5.1.** *Consider copula distributional members $C_i(u_1, u_2, \ldots, u_n) \in \mathcal{A}^n$, where $\mathcal{A}^n$ defines the space of all possible n-variate distributional members of the Archimedean family of copula models, specified in Lemma B.2. Any finite mixture distribution constructed from such copula components that admit tractable density functions $c_i(u_1, u_2, \ldots, u_n)$, denoted by*

$$\tilde{c}(u_1, u_2, \ldots, u_n) = \sum_{i=1}^{m} w_i c_i(u_1, u_2, \ldots, u_n),$$

*such that $\sum_{i=1}^{m} w_i = 1$, is also the density of a copula distribution.*

The proof of Lemma 5.1 is provided in Appendix C.

To proceed with developing a mixture copula structure for the PIC Bayesian model we first need to introduce the concept of data augmentation within a Bayesian model structure. This approach is typically invoked to deal with situations in which the likelihood evaluation would otherwise be intractable to evaluate point-wise. For instance 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 realisation of $\boldsymbol{Y}^{(1)}_{1:n}$ would require solving the integral in Eq. (5.1)

$$p\left(\boldsymbol{Y}^{(1)}_{1:n}|\boldsymbol{\theta}\right) = \int p\left(\boldsymbol{Y}^{(1)}_{1:n}|\boldsymbol{\theta}, \boldsymbol{Y}^{(2)}_{1:n}\right) p\left(\boldsymbol{Y}^{(2)}_{1:n}|\boldsymbol{\theta}\right) d\boldsymbol{Y}^{(2)}_{1:n}. \tag{5.1}$$

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 Eq. (5.2)

$$p\left(\boldsymbol{\theta}, \boldsymbol{Y}^{(2)*}_{1:n}|\boldsymbol{Y}^{(1)}_{1:n}\right) = \frac{p\left(\boldsymbol{Y}^{(1)}_{1:n}|\boldsymbol{\theta}, \boldsymbol{Y}^{(2)*}_{1:n}\right) p\left(\boldsymbol{Y}^{(2)*}_{1:n}|\boldsymbol{\theta}\right) p(\boldsymbol{\theta})}{p\left(\boldsymbol{Y}^{(1)}_{1:n}\right)} \tag{5.2}$$

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.

To interpret this approach within the PIC Bayesian model structure we first introduce some useful notation in Definition 5.2.

**Definition 5.2** (*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 $X = \left[\boldsymbol{X}_0', \boldsymbol{X}_1', \ldots, \boldsymbol{X}_J'\right]$. 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 $\widetilde{\boldsymbol{X}}_i = \mathcal{P}^*_{\boldsymbol{i}}(\boldsymbol{X}_i) = [\log P_{i,0}, \log P_{i,1}, \ldots, \log P_{i,J}, \log I_{i,0}, \log I_{i,1}, \ldots, \log I_{i,J-1}]$. 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 using Eq. (5.3) given in Box III. 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.

Given this notation we now observe that the PIC copula model equivalent of Eq. (5.2) is the observed data likelihood given for the $i$th accident year by

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

where matrix-variate Gaussian distributions $f^{MVN}_X()$ and $f^{MVN}_X$ are as defined in Lemma A.1 with $\widetilde{X}_{i,\text{aux}} = \text{Vec}\left(\widetilde{\boldsymbol{X}}^P_{i,\text{aux}}, \widetilde{\boldsymbol{X}}^I_{i,\text{aux}}\right)$, $M_{i,\text{aux}} = \text{Vec}\left([\Phi_{J-i+1:J}]', [\Psi_{J-i+1:J-1}]'\right)$, $\widetilde{X}_i = [\widetilde{\boldsymbol{X}}^P_{i,\text{obs}}, \widetilde{\boldsymbol{X}}^P_{i,\text{aux}}, \widetilde{\boldsymbol{X}}^I_{i,\text{obs}}, \widetilde{\boldsymbol{X}}^I_{i,\text{aux}}]$ and $M_i = [\Phi_0, \ldots, \Phi_J, \Psi_0, \ldots, \Psi_J]$ the equivalent mean. Clearly, the marginalisation required to evaluate the observed data likelihood involves intractable integration, except in special cases in which the copula models are Gaussian or independence copulas.

To overcome this intractability in evaluation of the likelihood we consider 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.

**Model Assumptions 5.3** (*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 $\widetilde{\boldsymbol{X}}_i = \left[\widetilde{\boldsymbol{X}}^P_{i,\text{obs}}, \widetilde{\boldsymbol{X}}^P_{i,\text{aux}}, \widetilde{\boldsymbol{X}}^I_{i,\text{obs}}, \widetilde{\boldsymbol{X}}^I_{i,\text{aux}}\right]$ with $\widetilde{\boldsymbol{X}}_i \in \mathbb{R}^{2J+1}$ for all $i = 0, \ldots, J$. We assume that $\Sigma$ is diagonal where*

$$\Sigma_{i,i} \sim \mathcal{IG}(\alpha_i, \beta_i), \quad \forall i \in \{0, \ldots, J\}, \tag{5.4}$$

  *where $\alpha_i$ and $\beta_i$ are the known hyper-parameters for shape and scale.*
- ***Marginal distribution**: given $\boldsymbol{\Theta} = \left(\Phi_0, \ldots, \Phi_J, \Psi_0, \ldots, \Psi_J\right)$ 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 ith accident year is matrix-variate Gaussian with density, defined as in Lemma A.1, with the $(2J+1) \times (J+1)$ mean matrix $\widetilde{M} = \left[\boldsymbol{\Theta}', \ldots, \boldsymbol{\Theta}'\right]$, column dependence given by $(2J+1) \times (2J+1)$ covariance matrix $\Sigma$ and row dependence given by $(J+1) \times (J+1)$ matrix $\Omega$. Here we only consider the case of $\Omega = \mathbb{I}_{J+1}$ for the marginal independent case.*
- ***Data augmented PIC copula likelihood**: Given $\widetilde{\boldsymbol{X}}^P_{0,\text{aux}}, \widetilde{\boldsymbol{X}}^P_{1,\text{aux}}, \ldots, \widetilde{\boldsymbol{X}}^P_{J-1,\text{aux}}, \widetilde{\boldsymbol{X}}^I_{0,\text{aux}}, \widetilde{\boldsymbol{X}}^I_{1,\text{aux}}, \ldots, \widetilde{\boldsymbol{X}}^I_{J-1,\text{aux}}$, $\boldsymbol{\Theta} = (\Phi_0, \ldots, \Phi_J, \Psi_0, \ldots, \Psi_J)$, 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 ith column (corresponding to ith accident year), the joint distribution of all losses $(\widetilde{\boldsymbol{X}}_i)$ is assumed to be hierarchical Archimedean Copula*

$$\begin{aligned}\widetilde{\boldsymbol{X}}_i &= \left[\widetilde{\boldsymbol{X}}^P_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^P_{i,\mathrm{aux}}, \widetilde{\boldsymbol{X}}^I_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^I_{i,\mathrm{aux}}\right] \\ &= \left[\widetilde{\boldsymbol{X}}^P_{0,i,\mathrm{obs}}, \ldots, \widetilde{\boldsymbol{X}}^P_{J-i,i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^P_{J-i+1,i,\mathrm{aux}}, \ldots, \widetilde{\boldsymbol{X}}^P_{J,i,\mathrm{aux}}, \widetilde{\boldsymbol{X}}^I_{0,i,\mathrm{obs}}, \ldots, \widetilde{\boldsymbol{X}}^I_{J-i,i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^I_{J-i+1,i,\mathrm{aux}}, \ldots, \widetilde{\boldsymbol{X}}^I_{J-1,i,\mathrm{aux}}\right] \\ &= \left[\underbrace{\log P_{i,0}, \ldots, \log P_{i,J-i}}_{\text{observed Payments}}, \underbrace{\log P_{i,J-i+1}, \ldots, \log P_{i,J}}_{\text{unobserved Payments}}, \underbrace{\log I_{i,0}, \ldots, \log I_{i,J-i}}_{\text{observed Incurred}}, \underbrace{\log I_{i,J-i+1}, \ldots, \log I_{i,J-1}}_{\text{unobserved Incurred}}\right]'. \end{aligned} \tag{5.3}$$

**Box III.**

*(HAC) mixture copula specified by distribution,*

$$\begin{aligned}\left[\widetilde{X}\right]_{\bullet,i} &\sim \widetilde{C}_{\boldsymbol{\rho}_i}\left(F\left(\widetilde{\boldsymbol{X}}^P_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^P_{i,\mathrm{aux}}, \widetilde{\boldsymbol{X}}^I_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^I_{i,\mathrm{aux}}; [\boldsymbol{M}]_{\bullet,i}, \Sigma\right)\right) \\ &= \widetilde{C}^P_{\boldsymbol{\rho}^P_i}\left(F\left(\widetilde{\boldsymbol{X}}^P_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^P_{i,\mathrm{aux}}; [\boldsymbol{M}]^P_{\bullet i}, \Sigma\right)\right) \\ &\quad \times \widetilde{C}^I_{\boldsymbol{\rho}^I_i}\left(F\left(\widetilde{\boldsymbol{X}}^I_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^I_{i,\mathrm{aux}}; [\boldsymbol{M}]^I_{\bullet i}, \Sigma\right)\right),\end{aligned} \tag{5.5}$$

*with super script $P$ and $I$ denote the components for the log payments and log incurred losses in the ith development year respectively and the density is given by*

$$\begin{aligned} &f\left(\widetilde{\boldsymbol{X}}^P_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^P_{i,\mathrm{aux}}, \widetilde{\boldsymbol{X}}^I_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^I_{i,\mathrm{aux}} | [\boldsymbol{M}]_{\bullet i}, \Sigma, \boldsymbol{\rho}^P_i, \boldsymbol{\rho}^I_i\right) \\ &\quad = \tilde{c}^P_{\boldsymbol{\rho}^P_i}\left(F\left(\widetilde{\boldsymbol{X}}^P_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^P_{i,\mathrm{aux}}; [\boldsymbol{M}]^P_{\bullet i}, \Sigma\right)\right) \\ &\qquad \times \tilde{c}^I_{\boldsymbol{\rho}_i}\left(F\left(\widetilde{\boldsymbol{X}}^I_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^I_{i,\mathrm{aux}}; [\boldsymbol{M}]^I_{\bullet i}, \Sigma\right)\right) \\ &\qquad \times \prod_{j=1}^{2J+1} \phi(\widetilde{X}_{j,i}; M_{j,i}, \Sigma_{i,i}),\end{aligned} \tag{5.6}$$

*where*

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

*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 Archimedean copula models made up of Frank, Clayton and Gumbel members, such that for the source of data S, the copula parameters for each Archimedean 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,*

$$\begin{aligned} &f\left(\tilde{X} | M, \Sigma, \Omega, \boldsymbol{\rho}\right) \\ &\quad = \underbrace{\prod_{i=0}^{J} \tilde{c}^P_{\boldsymbol{\rho}^P_i}\left(F\left(\widetilde{\boldsymbol{X}}^P_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^P_{i,\mathrm{aux}}; [\boldsymbol{M}]^P_{\bullet i}, \Sigma\right)\right) \tilde{c}^I_{\boldsymbol{\rho}^I_i}\left(F\left(\widetilde{\boldsymbol{X}}^I_{i,\mathrm{obs}}, \widetilde{\boldsymbol{X}}^I_{i,\mathrm{aux}}; [\boldsymbol{M}]^I_{\bullet i}, \Sigma\right)\right)}_{\text{Copula Dependence in Data Augmented PIC Likelihood}} \\ &\qquad \times \underbrace{\frac{\exp\left(-\frac{1}{2}\mathrm{tr}\left[\Omega^{-1}\left(\widetilde{X} - M\right)' \Sigma^{-1}\left(\widetilde{X} - M\right)\right]\right)}{(2\pi)^{(2J^2+3J+1)/2} |\Omega|^{(2J+1)/2} |\Sigma|^{(J+1)/2}}}_{\text{Marginal Distribution in Data Augmented Likelihood PIC Model}}.\end{aligned} \tag{5.7}$$

– *Assume that the tail dependence features of the Data-Augmented copula PIC model are such that the dependence structure is homogeneous across 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 $\Sigma$, $\boldsymbol{\Phi} = [\Phi_0, \Phi_1, \ldots, \Phi_J]$ and $\boldsymbol{\Psi} = [\Psi_0, \Psi_1, \ldots, \Psi_J]$ the hierarchical prior distribution on the auxiliary payment data for the ith accident year is given by a normal distribution, centred on the development year mean,*

$$\widetilde{\boldsymbol{X}}^P_{i,\mathrm{aux}} \sim \mathcal{N}\left(\left[\Phi_{J-i+1}, \Phi_{J-i+2}, \ldots, \Phi_J\right], \Sigma^P_2\right). \tag{5.8}$$

*The hierarchical prior distribution on the auxiliary incurred loss data for the ith accident year is given by*

$$\widetilde{\boldsymbol{X}}^I_{i,\mathrm{aux}} \sim \mathcal{N}\left(\left[\Psi_{J-i+1}, \Psi_{J-i+2}, \ldots, \Psi_J\right], \Sigma^I_2\right), \tag{5.9}$$

*with $\Sigma_2$ the lower portion of covariance $\Sigma$ 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 $\widetilde{\Sigma}$ is positive definite and components of $\boldsymbol{\Theta}$ are independent with prior distributions*

$$\Phi_i \sim \mathcal{N}\left(\phi_i, s_i^2\right) \quad \text{and} \quad \Psi_j \sim \mathcal{N}\left(\psi_j, t_j^2\right) \tag{5.10}$$

*and hyper-prior distributions*

$$s_i^2 \sim \mathcal{IG}\left(\alpha_i, \beta_i\right) \quad \text{and} \quad t_j^2 \sim \mathcal{IG}\left(a_j, b_j\right) \tag{5.11}$$

*for all $i \in \{1, \ldots, J\}$ and $j \in \{0, \ldots, J\}$.*
– *The matrix $\Sigma$ is distributed as $\Sigma \sim \mathcal{IW}(\Lambda, k)$ 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 auxiliary 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.

## 6. Efficient Markov chain Monte Carlo methods for PIC Bayesian models with dependence

Having defined a range of different Bayesian PIC models with different dependence structures we now develop a specialised framework for working with such models based on Markov chain Monte Carlo (MCMC) methods. This is important for actuaries who will require these methods to efficiently sample from the posterior distributions considered. This is particular the case in these models since standard MCMC methods may be very inefficient under such model structures.

We start by considering Gaussian copula models I, II and III and discuss a block-wise Metropolis within Gibbs sampler with specialised adaptive MCMC samplers for the covariance structures considered. In addition, we show that one can exploit for the posterior model parameters efficient full conditional conjugacy properties to provide several components which admit exact sampling in the block Gibbs MCMC structure.

### *6.1. Hierarchical Bayesian conjugacy under Gaussian 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 6.1.** *Conditional upon the parameters $\mathbf{\Theta}$ and the covariance matrix $\Sigma$, the permuted data $\mathcal{P}_{\boldsymbol{i}}^{*}(\text{Vec}(\boldsymbol{X}))$ can be transformed to produce the independent likelihood in Eq. (3.2). 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 $\Sigma = \mathbb{C}\text{ov}(\boldsymbol{X})$, then the resulting covariance structure $\mathbb{C}\text{ov}(\mathcal{T}(\boldsymbol{X})) = \mathbb{I}_d$. 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 $\Sigma$ and $\Lambda$ is a diagonal $d\times d$ matrix of the eigenvalues of $\Sigma$. Therefore the following holds for each of the models under a transform of the vector of permuted observations $\mathcal{T}\left(\mathcal{P}_{\boldsymbol{i}}^{*}(\text{Vec}(\boldsymbol{X}))\right)$:*

1. *Model II—When $\tilde{\Sigma} = \Sigma\otimes\Omega$, with $\Omega = \mathbb{I}_{J+1}$ then, $\mathcal{T}\left(\mathcal{P}_{\boldsymbol{i}}^{*}(\text{Vec}(\boldsymbol{X}))\right) = \left(U\Lambda^{\frac{1}{2}}\otimes\mathbb{I}_{J+1}\right)\mathcal{P}_{\boldsymbol{i}}^{*}(\text{Vec}(\boldsymbol{X}))$, where the $((2J+1)\times(2J+1))$ covariance $\Sigma$ is decomposed as $U\Lambda^{\frac{1}{2}}U'$.*
2. *Model II—When*
$$\tilde{\Sigma} = \bigoplus_{i=0}^{J}\Sigma_i,\ \mathcal{T}\left(\mathcal{P}_{\boldsymbol{i}}^{*}(\text{Vec}(\boldsymbol{X}))\right) = \left(\bigoplus_{i=0}^{J}U_i\Lambda_i^{\frac{1}{2}}\right)\mathcal{P}_{\boldsymbol{i}}^{*}(\text{Vec}(\boldsymbol{X})),$$
*where each accident year's dependence between payments and incurred losses is given by the $(2J+1)\times(2J+1)$ matrix $\Sigma_i$ which is decomposed as $U_i\Lambda_i^{\frac{1}{2}}U_i'$.*
3. *Model III—When $\tilde{\Sigma} = \left(\bigoplus_{i=0}^{J}\Sigma_0^{P}\right)\oplus\left(\bigoplus_{i=0}^{J}\Sigma_0^{I}\right)$,*
$$\mathcal{T}\left(\mathcal{P}_{\boldsymbol{i}}^{*}(\text{Vec}(\boldsymbol{X}))\right) = \left(\bigoplus_{i=0}^{J}U_i^{P}\left(\Lambda_i^{P}\right)^{\frac{1}{2}}\right) \oplus\left(\bigoplus_{i=0}^{J}U_i^{I}\left(\Lambda_i^{I}\right)^{\frac{1}{2}}\right)\mathcal{P}_{\boldsymbol{i}}^{*}(\text{Vec}(\boldsymbol{X}))$$
*where each of the covariance matrices $\Sigma_i^{P}$ and $\Sigma_i^{I}$ 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 $\mathcal{T}\left(\mathcal{P}_{\boldsymbol{i}}^{*}(\text{Vec}(\boldsymbol{X}))\right)$, 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 Wüthrich (2010) as defined in Eq. (3.2). Of course this is defined now with respect to components in the likelihood corresponding to the transformed components, as detailed in Eq. (4.11).*

**Remark 6.2.** The consequence is that results in Lemma 6.1 are that the conjugacy properties derived for the independent model in Merz and Wüthrich (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 6.3.** *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 $\mathcal{T}\left(\mathcal{P}_{\boldsymbol{i}}^{*}(\text{Vec}(\boldsymbol{X}))\right)$ on the data, described in Lemma 6.1, 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 Wüthrich, 2010, Theorem 3.4 for the independent case):*

$$\left[\widetilde{\Phi}_{0:J},\widetilde{\Psi}_{0:J}\,|\,\Sigma,\Omega,\mathcal{T}\left(\mathcal{P}_{\boldsymbol{i}}^{*}(\text{Vec}(\boldsymbol{X}))\right)\right]\sim\mathcal{N}(\Pi,\Delta) \tag{6.1}$$

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

$$\begin{aligned} a_{n,m} &= \left(s_n^{-2}+(J-n+1)\sigma_n^{-2}\right)\delta_{n=m} \\ &\quad + \sum_{i=0}^{(n-1)\wedge(m-1)}\left(\nu_i^2-\omega_i^2\right)^{-1}, \quad \text{for } 0\le n,m\le J, \\ a_{J+1+n,J+1+m} &= \left(t_n^{-2}+(J-n)\tau_n^{-2}\right)\delta_{n=m} \\ &\quad + \sum_{i=0}^{n\wedge m}\left(\nu_i^2-\omega_i^2\right)^{-1}, \end{aligned} \tag{6.2}$$

*for $0\le n,m\le J-1$,*

$$a_{n,J+1+m} = \Delta_{n,J+1+m} = -\sum_{i=0}^{(n-1)\wedge m}\left(\nu_i^2-\omega_i^2\right)^{-1},$$

*for $0\le n\le J,\ 0\le m\le J-1$;*

*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,*

$$\left[\widetilde{\Phi}_{0:J},\widetilde{\Psi}_{0:J}\right] = \Delta\left(\widetilde{c}_0,\widetilde{c}_1,\ldots,\widetilde{c}_J,\widetilde{b}_0,\ldots,\widetilde{b}_J\right), \tag{6.3}$$

*with*

$$\begin{aligned} \widetilde{c}_j &= s_j^{-2}\phi_j+\sigma_j^{2}\sum_{i=0}^{J-j}\log\left(\frac{\widetilde{P}_{i,j}}{\widetilde{P}_{i,j-1}}\right) \\ &\quad + \sum_{i=J-j+1}^{J}\left(\nu_{J-i}^2-\omega_{J-i}^2\right)^{-1}\log\left(\frac{\widetilde{I}_{i,J-i}}{\widetilde{P}_{i,J-i}}\right), \\ \widetilde{b}_j &= t_j^{-2}\psi_j+\tau_j^{2}\sum_{i=0}^{J-j-1}\log\left(\frac{\widetilde{I}_{i,j}}{\widetilde{I}_{i,j+1}}\right) \\ &\quad - \sum_{i=J-j}^{J}\left(\nu_{J-i}^2-\omega_{J-i}^2\right)^{-1}\log\left(\frac{\widetilde{I}_{i,J-i}}{\widetilde{P}_{i,J-i}}\right). \end{aligned} \tag{6.4}$$

*Given the transform vector $\left[\widetilde{\Phi}_{0:J},\widetilde{\Psi}_{0:J}\right]$, the parameters on the original 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*
$$\left[\Phi_{0:J},\Psi_{0:J}\right]' = U^{-1}\Lambda^{-\frac{1}{2}}\left[\widetilde{\Phi}_{0:J},\widetilde{\Psi}_{0:J}\right]. \tag{6.5}$$
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,*
$$\left[\Phi_{0:J},\Psi_{0:J}\right]' = U_i^{-1}\Lambda_i^{-\frac{1}{2}}\left[\widetilde{\Phi}_{0:J},\widetilde{\Psi}_{0:J}\right]. \tag{6.6}$$
3. *Model III—On the untransformed scale, the solution is given by the following system of equations,*
$$\left[\Phi_{0:J},\Phi_{0:J-1},\Phi_{0:J-2},\ldots,\Phi_J\right]'$$

$$= \bigoplus_{i=0}^{J} \left(U_i^P\right)^{-1} \left(\Lambda_i^P\right)^{-\frac{1}{2}} \left[\widetilde{\Phi}_{0:J}, \widetilde{\Phi}_{0:J-1}, \widetilde{\Phi}_{0:J-2}, \ldots, \widetilde{\Phi}_J\right],$$

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

- ***Conjugate posterior distribution for the covariance matrix**: Given the transformed observed payment and incurred losses have a multivariate Gaussian likelihood, as specified in Eq. (4.21), with covariance matrix $\widetilde{\Sigma} = \Sigma \otimes \Omega$ and mean vector $\mathrm{Vec}(M)$. Then the posterior for the covariance matrix is the Inverse-Wishart–Gaussian distribution detailed in* Kannan et al. (2010, Section 3) *and Kannan et al. (2011)*

$$\left[\widetilde{\Sigma} | \Phi_{0:J}, \Psi_{0:J}, \mathcal{T}\left(\mathcal{P}_{\boldsymbol{i}}^{*}\left(\mathrm{Vec}(\boldsymbol{X})\right)\right)\right] \sim \mathcal{IW}\Big(\Lambda + \mathcal{T}\left(\mathcal{P}_{\boldsymbol{i}}^{*}\left(\mathrm{Vec}(\boldsymbol{X})\right)\right) \mathcal{T}\left(\mathcal{P}_{\boldsymbol{i}}^{*}\left(\mathrm{Vec}(\boldsymbol{X})\right)\right)', \dim\left(\mathrm{Vec}(\boldsymbol{X})\right) + \widetilde{k}\Big). \tag{6.7}$$

*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 A.2 *and the result in Eq.* (6.7) *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,*

$$\left[s_i^2 | \Phi_i\right] \sim \mathcal{IG}\left(\alpha_i + \frac{1}{2}, \beta_i + \frac{(\Phi_i - \phi_i)^2}{2}\right) \quad \text{and}$$

$$\left[t_i^2 | \Psi_i\right] \sim \mathcal{IG}\left(a_i + \frac{1}{2}, b_i + \frac{(\Psi_i - \psi_i)^2}{2}\right).$$

### 6.2. Estimation via adaptive data-augmented MCMC for claims reserving 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.

It is not familiar to actuarial literature that there are now 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 utilise 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. (2005) 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 $Q_\gamma$, 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 $Q_\gamma$ 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, $\Gamma_s$ 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*: $\lim_{j\to\infty} \|\mathcal{L}\mathrm{aw}(\boldsymbol{\theta}) - P(\boldsymbol{\theta})\|_{tv} = 0$ 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 paper 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 utilising 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 Adaptive Metropolis algorithm that we utilise is combined 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

$$\boldsymbol{\Upsilon} = \left[\boldsymbol{\Phi}, \boldsymbol{\Psi}, s_{0:J}^2, t_{0:J}^2, \boldsymbol{\rho}, \widetilde{\boldsymbol{X}}_{1,\mathrm{aux}}^P, \ldots, \widetilde{\boldsymbol{X}}_{J,\mathrm{aux}}^P, \widetilde{\boldsymbol{X}}_{1,\mathrm{aux}}^I \ldots, \widetilde{\boldsymbol{X}}_{J,\mathrm{aux}}^I\right]$$

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. We provide the specific details of the adaptive MCMC proposal structures we recommend for adaptively proposing MCMC moves in both parameters with posterior support in Euclidean space and posterior parameters with support on the manifold for positive definite symmetric matrices, see Appendix D.

## 7. Analysis of MCMC sampler efficiency: real data illustration

To illustrate the proposed models and compare with existing models and estimation methods in the actuarial literature we consider, as in Merz and Wüthrich (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 Wüthrich (2010), we treat the claim development factors, the likelihood dependence parameters and the hyper-parameters on the claim development factor priors as parameters which we incorporate into the posterior inference.

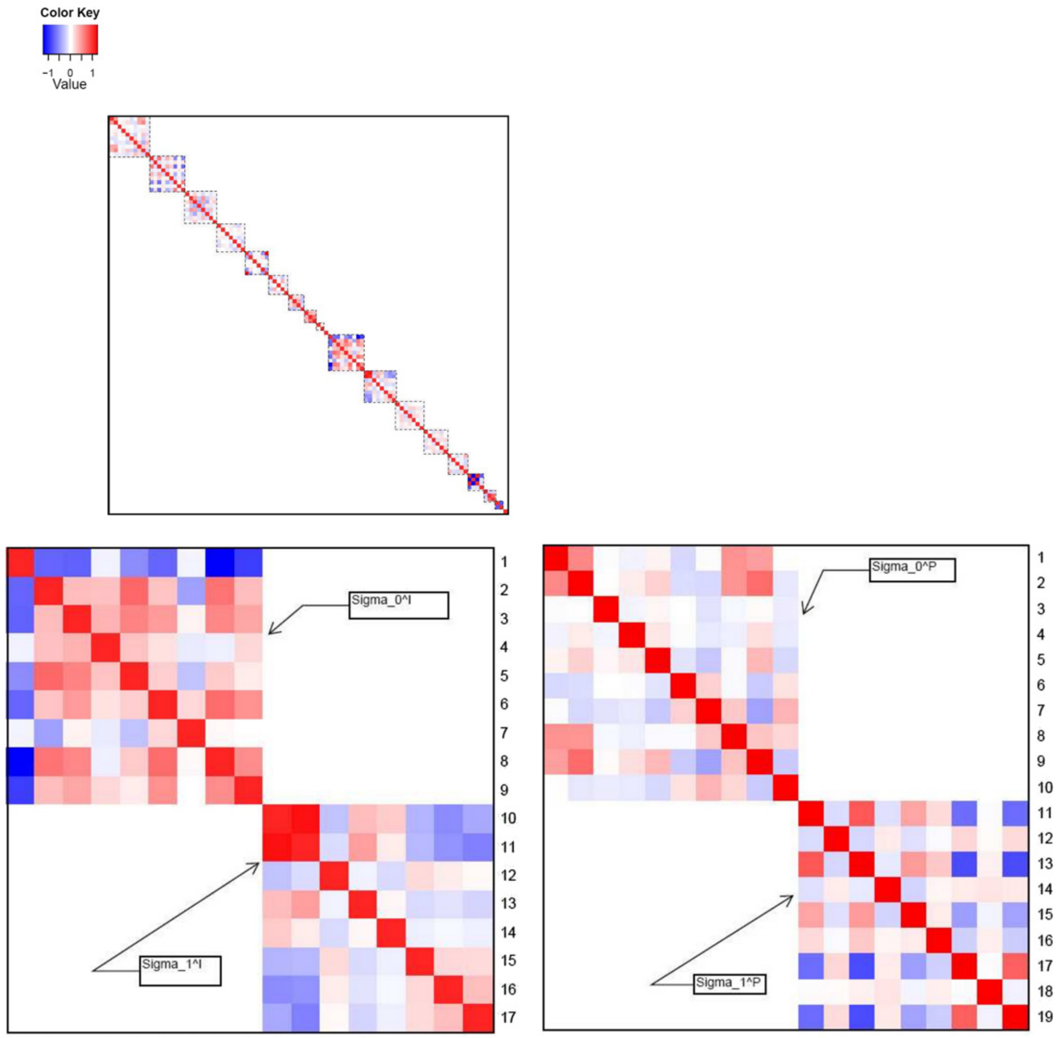


**Fig. 2.** *Top panel:* Heatmap of the posterior distribution for the Gaussian copula covariance matrix (100 × 100), summarised by the heat map for the mean of correlation structure using samples from the Riemannian Manifold Adaptive Metropolis sampler under restriction to a telescoping diagonal block form. *Bottom Left Panel*: Heatmap for the posterior distribution sub-block covariance matrices $\Sigma_0^P$ and $\Sigma_1^P$ converted to correlation matrices. *Bottom Right Panel*: Heatmap for the posterior distribution sub-block covariance matrices $\Sigma_0^I$ and $\Sigma_1^I$ converted to correlation matrices. The colour key is given at the top left.

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 are relevant for practical outputs for actuarial applications which involves assessing 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 the simulation 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 Eq. (D.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 4.2, Eq. (4.23). 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.

**Convergence analysis**: In all the Markov chain Monte Carlo simulations, for each model (payment, payment–incurred Gaussian copula Model III; and Data-Augmented hierarchical Archimedean 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 burn in and the remaining samples were used in inference results presented below.

### *7.1. Hierarchical Bayesian PIC model III: adaptive MCMC results*

This section presents the estimation results for the Gaussian Copula based PIC models (Model III) on the real data. Fig. 2 sum-

**Table 1**
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 eigenvector $\widehat{\boldsymbol{v}}$ |
|---|---|---|---|---|
| $\Sigma_0^P$ | 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] |
| $\Sigma_1^P$ | 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] |
| $\Sigma_2^P$ | 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] |
| $\Sigma_3^P$ | 0.75 | 6.38 | [0.14; 1.92] | [0.08, −0.12, −0.05, −0.03, −0.01, −6.9e−5, −2.3e−3] |
| $\Sigma_4^P$ | 0.76 | 6.81 | [0.13; 0.25] | [0.12, −0.13, −0.06, −0.03, −0.01, 0.01] |
| $\Sigma_5^P$ | 0.70 | 5.93 | [0.12; 0.23] | [0.14, −0.15, −0.08, −0.01, 0.01] |
| $\Sigma_6^P$ | 1.11 | 9.90 | [0.12; 0.24] | [0.21, −0.20, −0.07, 0.03] |
| $\Sigma_7^P$ | 2.16 | 18.07 | [0.10; 0.26] | [0.27, −0.25, 0.07] |
| $\Sigma_8^P$ | 5.44 | 34.67 | [0.08; 20.92] | [−0.47, 0.43] |
| $\Sigma_9^P$ | 1.95 | 10.28 | [0.06; 11.56] | Not applicable |
| $\Sigma_0^I$ | 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] |
| $\Sigma_1^I$ | 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] |
| $\Sigma_2^I$ | 0.80 | 3.26 | [0.12; 3.76] | [0.09, −0.12, −0.06, −0.02, −0.01, 3.2e−3, −0.01] |
| $\Sigma_3^I$ | 0.66 | 3.15 | [0.11; 3.27] | [0.10, −0.12, −0.07, −0.03, −0.02, 1.5e−3] |
| $\Sigma_4^I$ | 0.65 | 3.97 | [0.10; 3.23] | [0.15, −0.15, −0.07, −0.02, 0.02, ] |
| $\Sigma_5^I$ | 1.00 | 5.02 | [0.09; 5.62] | [6.6e−12, −6.6e−12, −2.9e−12, −1.0e−12, 8.5e−13] |
| $\Sigma_6^I$ | 1.15 | 7.52 | [0.08; 6.31] | [0.31, −0.24, 0.07] |
| $\Sigma_7^I$ | 5.26 | 25.29 | [0.06; 32.87] | [−0.50, 0.42] |
| $\Sigma_8^I$ | 1.03 | 3.82 | [0.04; 6.30] | Not applicable |

marises the dependence structure by a heatmap for the posterior distribution of the Gaussian copula covariance matrix. 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,

$$\mathbb{E}\left[\tilde{\Sigma}|\boldsymbol{P},\boldsymbol{I}\right]=\frac{1}{S}\sum_{s=1}^{S}\left\{\left(\bigoplus_{i=0}^{J}\Sigma_i^P\right)\oplus\left(\bigoplus_{i=0}^{J}\Sigma_i^I\right)\right\}^{(s)}, \tag{7.1}$$

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 Fig. 2. 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 1 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 payment and incurred data as summary statistics. These estimates are given by

$$\widehat{\lambda}_i^{(s)}=\arg\max\left(\det(\Sigma_i^{(s)}-\lambda\mathbb{I})=0\right). \tag{7.2}$$

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 across 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 1 summarises the results for the average PCA weight (largest eigenvalue) and average posterior eigenvector.

Tokuda et al. (2011) develops a framework which formalises 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 Fig. 3. This involves the presentation of contour maps of these marginal posteriors that are constructed using adaptive MCMC samples of these matrices.

In Fig. 4, the development factors for payment and incurred data marginal posterior distributions are presented along with the posteriors of the hyper-parameters for the Gaussian Copula based PIC models (Model III). 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)

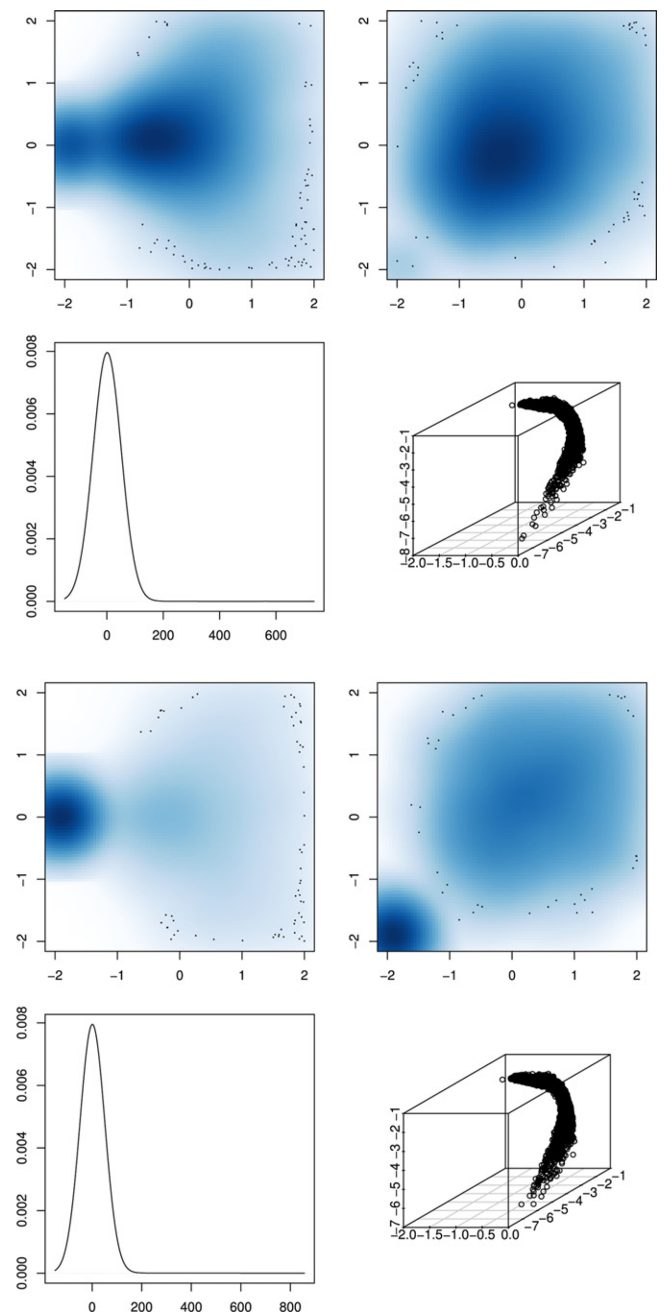

**Fig. 3.** Heatmaps for the block diagonal covariance matrices $\Sigma_4^P$ ($2 \times 2$ sub-plot 1) and $\Sigma_4^I$ ($2 \times 2$ sub-plot 2). These are obtained using samples from the Riemannian Manifold Adaptive Metropolis sampler. Samples from the Posterior distribution of the telescoping diagonal 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 $4 \times 4$ panels, starting from the top, summarises 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 $\left\{\log \operatorname{tr}\left(\Sigma_4^s\right)\right\}$. *Bottom Right Panel*: scatter plot of posterior samples of the first, second and third largest eigenvalues scaled by total of the eigenvalues (PCA weights—for linear combinations of the development factors when explaining variation in observed payment and incurred data for a given accident year).

dependent model; the PIC [Full] independent model and the PIC [Partial] independent model of Merz and Wüthrich (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 2.

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.

### 7.2. Data-augmented hierarchical Bayesian PIC model IV: adaptive MCMC results

This section presents the estimation results for the mixture of Clayton and Gumbel Copula based PIC models (Model IV) on the real data. Fig. 5 presents a summary of the mixture copula dependence structure obtained from posterior samples of the copula parameters under the hierarchical Bayesian model. 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 estimated 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 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.

Fig. 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).

## 8. Bayesian PIC copula reserving estimates

This section studies the effect on reserving estimates obtained by incorporating dependence structures into the PIC model. 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

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

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

In general it is not possible to solve these integrals analytically. However, for the Gaussian copula models developed in this paper, under the results in Lemma 6.1, one adopt the results of Merz and Wüthrich (2010, Theorem 2.4) to obtain analytic Gaussian predictive distributions. Alternatively, the predictive distributions can be estimated as described in Peters et al. (2010, Section 3.3). Although the results in Table 2 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 Archimedean copula model, it does not admit

**Table 2**
Posterior marginal distributions for development factors.

| Factor | PIC Gaussian copula (Full) | | PIC Independent (Full) | | Payment or incurred only (Full) | | Merz and Wüthrich (2010) PIC Independent (Partial) | PIC Mixture Clayton–Gumbel copula (Full) | |
|---|---|---|---|---|---|---|---|---|---|
| | Post. Ave. | [Q0.05; Q0.95] | Post. Ave. | [Q0.05; Q0.95] | Post. Ave. | [Q0.05; Q0.95] | Post. Ave. | Post. Ave. | [Q0.05; Q0.95] |
| $\Phi_0$ | 13.79 | [13.55; 14.03] | 14.51 | [13.19; 15.01] | 13.77 | [13.68; 13.86] | 13.78 | 13.72 | [13.63; 15.96] |
| $\Phi_1$ | 0.21 | [−0.16; 0.58] | 0.18 | [0.05; 0.29] | 0.20 | [0.12; 0.27] | 0.22 | 13.98 | [13.64; 15.05] |
| $\Phi_2$ | 0.25 | [−0.25; 0.77] | 0.22 | [0.08; 0.34] | 0.23 | [0.14; 0.31] | 0.24 | 14.10 | [12.80; 15.14] |
| $\Phi_3$ | 0.18 | [−0.44; 0.81] | 0.17 | [0.04; 0.30] | 0.15 | [0.06; 0.24] | 0.17 | 14.32 | [13.92; 15.29] |
| $\Phi_4$ | 0.15 | [−0.55; 0.86] | 0.16 | [0.02; 0.30] | 0.13 | [0.04; 0.23] | 0.16 | 14.61 | [14.32; 15.60] |
| $\Phi_5$ | 0.13 | [−0.63; 0.91] | 0.15 | [1.9e−3; 0.30] | 0.12 | [0.01; 0.22] | 0.14 | 14.71 | [14.64; 16.85] |
| $\Phi_6$ | 0.10 | [−0.71; 0.92] | 0.12 | [−0.04; 0.30] | 0.08 | [−0.04; 0.20] | 0.11 | 14.92 | [14.82; 16.50] |
| $\Phi_7$ | 0.07 | [−0.79; 0.93] | 0.13 | [−0.05; 0.33] | 0.05 | [−0.09; 0.19] | 0.07 | 14.96 | [14.90; 16.44] |
| $\Phi_8$ | 0.08 | [−0.81; 0.97] | 0.11 | [−0.09; 0.32] | 0.05 | [−0.12; 0.22] | 0.05 | 15.06 | [14.97; 17.59] |
| $\Phi_9$ | 0.04 | [−0.88; 0.98] | 0.10 | [−0.04; 0.52] | 0.02 | [−0.19; 0.24] | 0.08 | 15.10 | [13.26; 15.86] |
| $\Psi_0$ | 0.51 | [−0.84; 1.85] | 0.45 | [0.31; 0.56] | 0.52 | [0.38; 0.64] | 0.50 | 13.73 | [13.69; 15.33] |
| $\Psi_1$ | −0.15 | [−1.50; 1.20] | −0.08 | [−0.11; 0.12] | 0.01 | [−0.11; 0.12] | −0.15 | 14.01 | [13.94; 15.78] |
| $\Psi_2$ | −0.13 | [−1.49; 1.23] | −0.09 | [−0.15; 0.20] | 0.01 | [−0.12; 0.12] | −0.14 | 14.30 | [14.23; 15.77] |
| $\Psi_3$ | −3.7e−2 | [−1.39; 1.34] | 0.01 | [−0.05; 0.21] | 0.01 | [−0.13; 0.13] | −0.04 | 14.54 | [14.43; 16.46] |
| $\Psi_4$ | −1.7e−2 | [−1.39; 1.36] | −0.01 | [−0.06; 0.23] | −0.01 | [−0.15; 0.14] | −0.02 | 14.67 | [14.58; 16.07] |
| $\Psi_5$ | −7.1e−3 | [−1.39; 1.38] | 0.02 | [−0.06; 0.21] | −0.06 | [−0.17; 0.15] | −0.02 | 14.89 | [14.72; 18.10] |
| $\Psi_6$ | −7.3e−3 | [−1.40; 1.39] | −0.02 | [−0.05; 0.30] | −0.01 | [−0.19; 0.16] | −0.01 | 14.82 | [14.51; 15.87] |
| $\Psi_7$ | −2.4e−3 | [−1.40; 1.39] | 0.02 | [−0.05; 0.34] | −0.06 | [−0.40; 0.22] | −0.01 | 14.85 | [14.62; 16.05] |
| $\Psi_8$ | −2.0e−4 | [−1.40; 1.40] | −0.01 | [−0.02; 0.52] | −0.13 | [−0.52; 0.25] | −0.01 | 15.07 | [14.98; 16.76] |

*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 Wüthrich (2010) are the Bayesian posterior results in which $\sigma_m$ and $\tau_n$ 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).

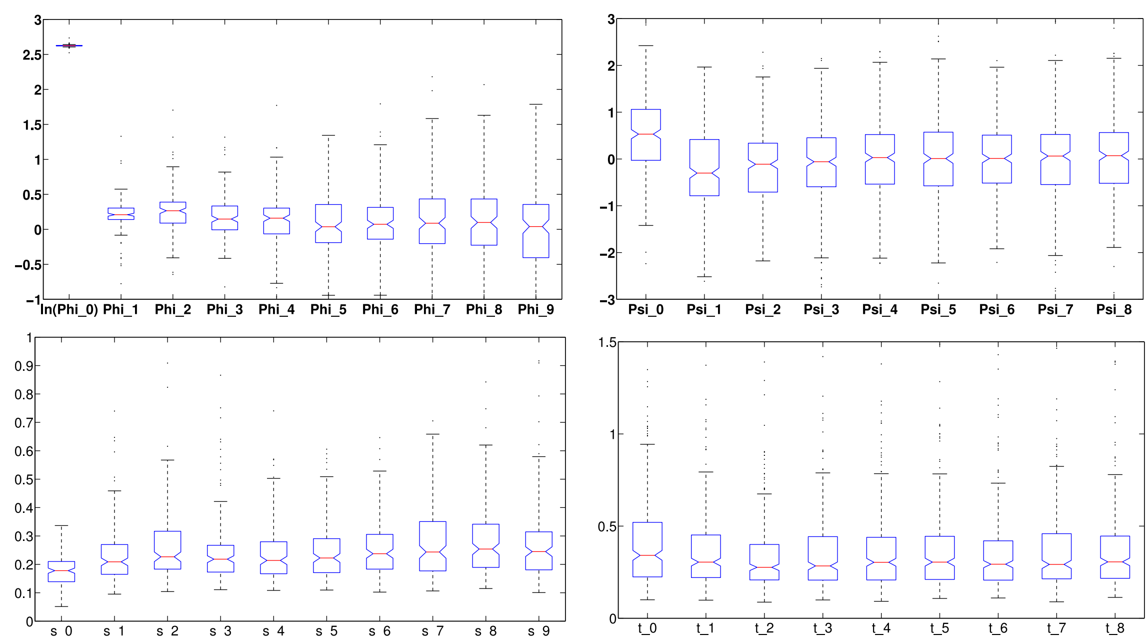


**Fig. 4.** Boxplot summaries of the marginal posterior distributions obtained using samples from the Riemannian Manifold Adaptive Metropolis sampler. Samples from the Posterior distribution under a telescoping diagonal block size form covariance matrix structures of the Gaussian copula under the hierarchical 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 $p(\Phi_i|\boldsymbol{P},\boldsymbol{I})$. *Top Right Panel*: box plots of marginal posterior distributions for $p(\Psi_i|\boldsymbol{P},\boldsymbol{I})$. *Bottom Left Panel*: box plots of marginal posterior distributions for $p(s_i|\boldsymbol{P},\boldsymbol{I})$. *Bottom Right Panel*: box plots of marginal posterior distributions for $p(t_i|\boldsymbol{P},\boldsymbol{I})$.

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.

Fig. 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 Bayesian models which incorporate priors on observation error, development factors and hyperpriors for precision of the development factors. 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.

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. Fig. 8 presents the MCMC estimated claims reserve marginal posterior predictive distribu-

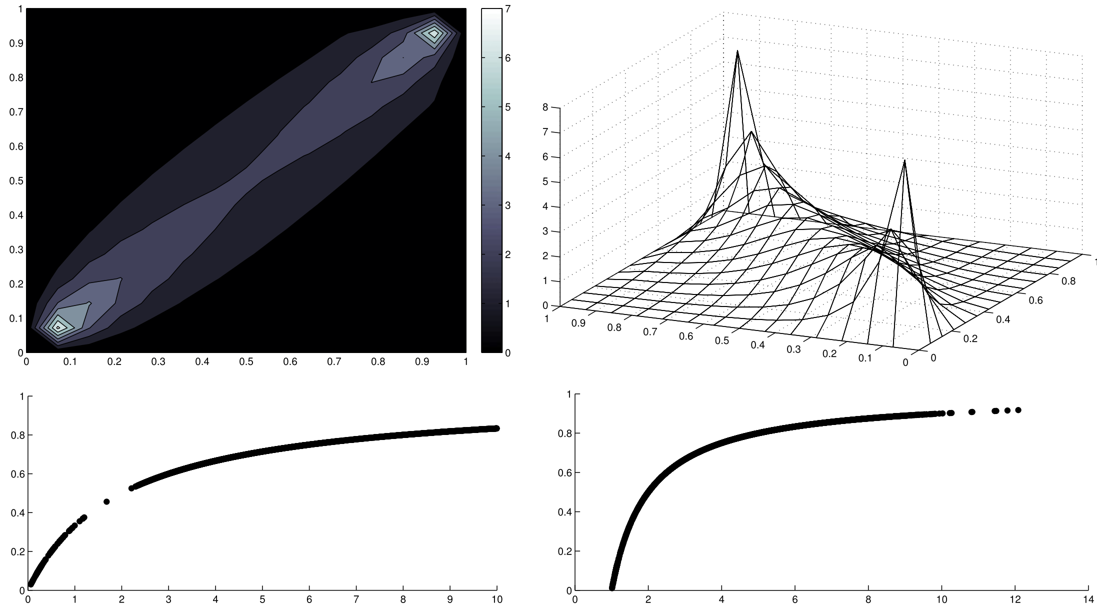

**Fig. 5.** Copula Dependence Parameter Posterior distributions estimated under the Data-Augmented Mixture Copula PIC Model IV. A mixture of Archimedean copula models is considered, with Clayton and Gumbel copula choices, allow for possible asymmetry in the tail dependence over development years. We 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 $\rho_C^{\mathrm{MMSE}}$ and $\rho_G^{\mathrm{MMSE}}$). *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 $\rho_C^{\mathrm{MMSE}}$ and $\rho_G^{\mathrm{MMSE}}$). *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.

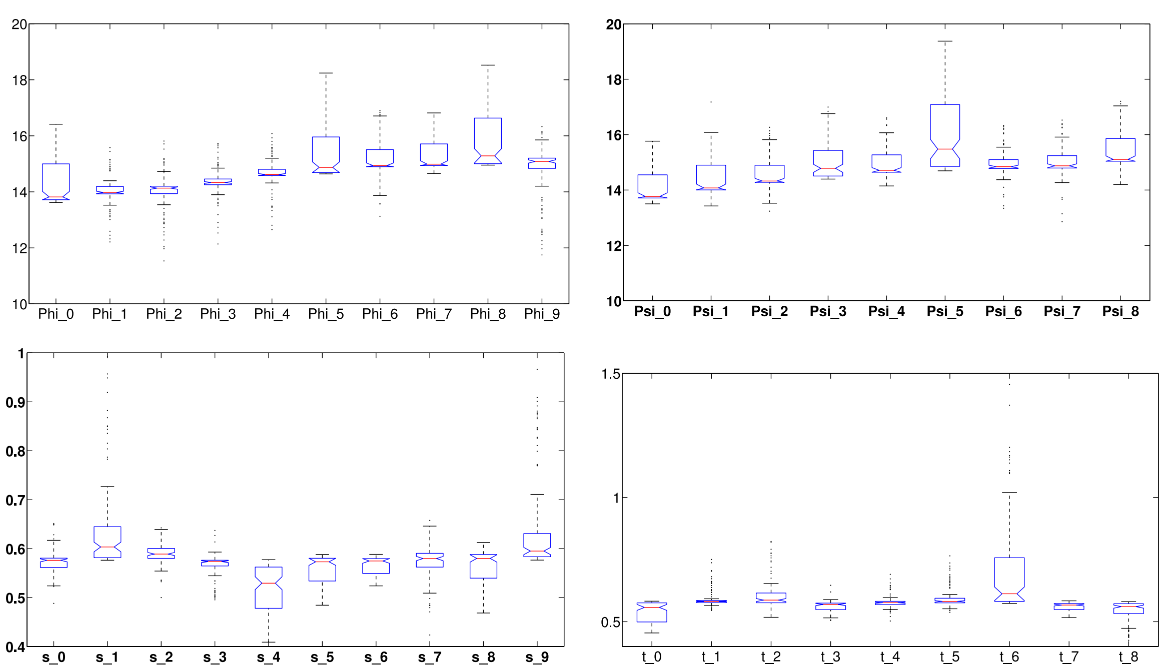


**Fig. 6.** Boxplots of the marginal posterior distributions of the development factors and hyperparameters. *Top Left Panel*: box plots of marginal posterior distributions for $p(\Phi_i|\boldsymbol{P}, \boldsymbol{I})$. *Top Right Panel*: box plots of marginal posterior distributions for $p(\Psi_i|\boldsymbol{P}, \boldsymbol{I})$. *Bottom Left Panel*: box plots of marginal posterior distributions for $p(s_i|\boldsymbol{P}, \boldsymbol{I})$. *Bottom Right Panel*: box plots of marginal posterior distributions for $p(t_i|\boldsymbol{P}, \boldsymbol{I})$.

tions for each accident year per model developed. We compared our results to those obtained in Merz and Wüthrich (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 literature can be obtained for the models of Mack (1993), Dahms (2008) and Quarg and Mack (2004) for this data analysis in Merz and Wüthrich (2010, Table 4) and in the spreadsheet provided by Professor Mario Wüthrich at URL.[1]

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

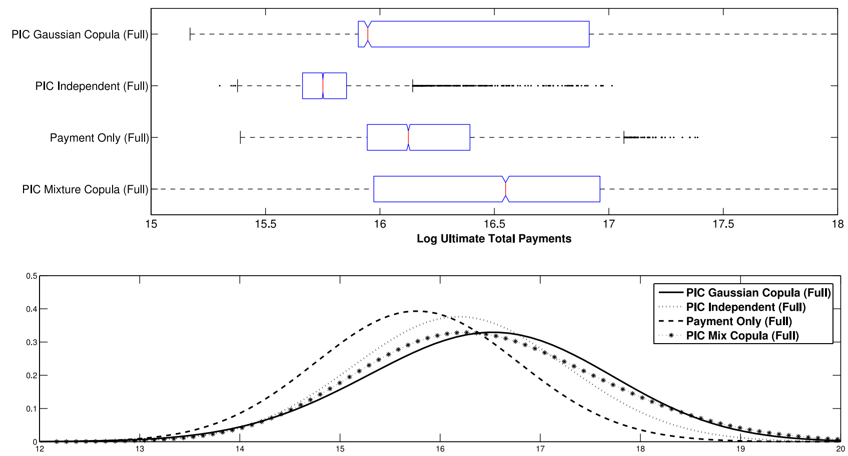


**Fig. 7.** Boxplots of the predictive distributions obtained from the MCMC samples. Ultimate Bayesian predictive distributions for log payment data from the payment only predictive distribution, the Full Independent PIC model, and the hierarchical PIC Mixture Copula model via Data Augmentation predictive distribution. *Left Panel*: Posterior predictive distribution box plots from samples. *Right Panel*: Kernel density estimates of the predictive distributions.

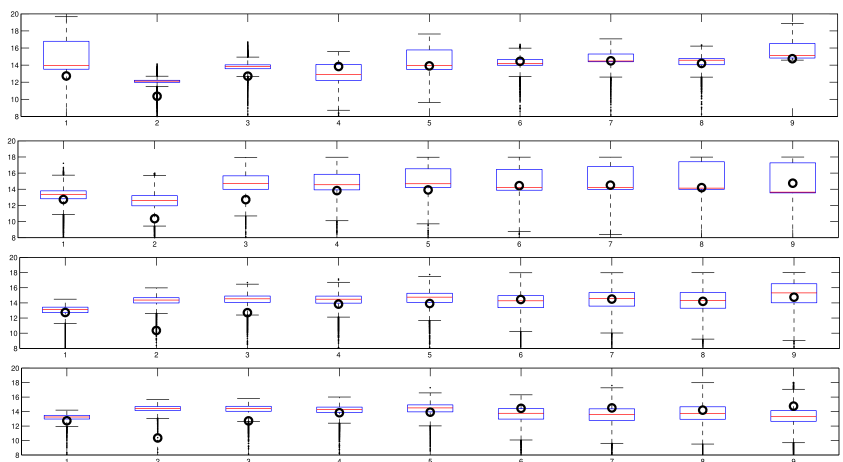

**Fig. 8.** Boxplots of log ultimate Bayesian predictive reserve distributions for payment data per accident year, compared to (Partial) PIC Independent posterior mean estimates from Merz and Wüthrich (2010) (large unfilled black circles). *Top Row*: the (Full) hierarchical PIC Mixture Copula model via 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.

## 9. Conclusions

This paper extends the class of PIC models to combine the two different channels of information as proposed in Merz and Wüthrich (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. These included generalised 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 Wüthrich (2013). Second, we develop a class of hierarchical mixture Archimedean 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 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 (positive 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.

## Acknowledgements

We would like to thank Prof. Mario Wüthrich 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.

## Appendix A

**Lemma A.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* $\Sigma$ *and* $\Psi$ *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* $\mathrm{Vec}\left(X'\right)$ *has a multivariate normal* *distribution* $\mathrm{Vec}\left(X'\right) \sim N\left(\mathrm{Vec}(M'), \Sigma \otimes \Psi\right)$. *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*

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

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

1. *If* $X \sim N_{p,n}(M, \Sigma \otimes \Psi)$, *then* $X' \sim N_{n,p}\left(M', \Psi \otimes \Sigma\right)$.
2. *If* $X \sim N_{p,n}(M, \Sigma \otimes \Psi)$, *and partition* $X$, $M$, $\Sigma$, *and* $\Psi$ *as*

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

*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*

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

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

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

*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 A.2** (*Properties of Matrix Variate Wishart Distributions*)**.** *Suppose that the* $(p \times p)$ *positive definite matrix* $\Sigma$ *has an inverse Wishart distribution, with positive definite* $(p \times p)$ *scale matrix* $\Lambda$, *degrees of freedom parameter* $k > p - 1$, *and density*

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

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

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

*Furthermore, the following marginal and conditional properties of the inverse Wishart distribution are relevant. Consider a partition of the matrices* $\Lambda$ *and* $\Psi$ *as*

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

*with* $\Lambda_{ij}$ *and* $\Sigma_{ij}$ *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* $\Sigma_{11}$ *is independent of* $\Sigma_{11}^{-1}\Sigma_{12}$.
2. *The marginal distribution of any sub matrix on the diagonal of the matrix* $\Sigma$ *is distributed as inverse Wishart. For example, the sub random matrix* $\Sigma_{11}$ *is as inverse Wishart with* $\Sigma_{11} \sim \mathcal{IW}(\Lambda_{11}, k - p_2)$;
3. *The marginal distribution of sub random matrix* $\Sigma_{22\cdot 1}$ *is inverse Wishart* $\Sigma_{22\cdot 1} \sim \mathcal{IW}(\Lambda_{22\cdot 1}, k)$.

In Lemma A.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 A.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* $\Psi$, *denoted* $V \sim IW_p(m, \Psi)$ *with density*

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

## Appendix B

The family of Archimedean copula models has the following useful properties presented in Lemma B.1.

**Lemma B.1.** *Let* $C$ *be an Archimedean copula with generator* $\varphi$. *Then according to* Nelsen (2006, Lemma 4.1.2 and Theorem 4.1.5)*, the following properties hold:*

1. *C is an Archimedean copula if it can be represented by*
$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$$
*where* $\varphi$ *is the generator of this copula and is a continuous, strictly decreasing function from* $[0, 1]$ *to* $[0, \infty]$ *such that* $\varphi(1) = 0$ *and* $\varphi^{[-1]}$ *is the pseudo inverse of* $\varphi$.
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* $\varphi$ *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 B.2.

**Lemma B.2.** *From the results in* Nelsen (2006, Section 4.3, Table 4.1) *the distribution and density functions of the Clayton copula are given respectively as:*

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

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

*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*

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

*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*

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

*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*

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

## Appendix C

**Proof.** The proof of Lemma 5.1 requires one to demonstrate that the resulting distribution function

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

satisfies the two conditions of a $n$-variate copula distribution given in (Definition 2.10.6) of Nelsen (2006). 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 $\widetilde{C}(\boldsymbol{u}) = 0$ if at least one coordinate of $\boldsymbol{u}$ is 0. Clearly since we have shown that $\widetilde{C}(\boldsymbol{u}) = \sum_{i=1^m} w_i C_i(\boldsymbol{u})$ 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 $\widetilde{C}$ is satisfied, $V_{\widetilde{C}}([\boldsymbol{a}, \boldsymbol{b}]) \geq 0$. As in Nelsen (2006) 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 $\widetilde{C}$ by (Definition 2.10.1, p. 43) of Nelsen (2006) giving

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

where the domain $\operatorname{Dom}\widetilde{C}$ of the mixture copula $\widetilde{C}$ satisfies $[\boldsymbol{a}, \boldsymbol{b}] \subseteq \operatorname{Dom}\widetilde{C}$. 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 $\operatorname{sgn}(\boldsymbol{c}) = 1$ if $c_k = a_k$ for an even number of $k$'s or $\operatorname{sgn}(\boldsymbol{c}) = -1$ if $c_k = a_k$ for an odd number of $k$'s. Equivalently, we consider

$$\begin{aligned}
\triangle_{a_k}^{b_k} \widetilde{C}(\boldsymbol{t}) &= \widetilde{C}(t_1, t_2, \ldots, t_{k-1}, b_k, t_{k+1}, \ldots, t_n) \\
&\quad - \widetilde{C}(t_1, t_2, \ldots, t_{k-1}, a_k, t_{k+1}, \ldots, t_n).
\end{aligned}$$

In the case of the mixture copula, we can expand the volume of the $n$-box $[\boldsymbol{a}, \boldsymbol{b}]$ as follows

$$\begin{aligned}
V_{\widetilde{C}}([\boldsymbol{a}, \boldsymbol{b}]) &= \sum \operatorname{sgn}(\boldsymbol{c}) \widetilde{C}(\boldsymbol{c}) = \sum_{i=1}^{m} \sum w_i \operatorname{sgn}(\boldsymbol{c}) C_i(\boldsymbol{c}) \\
&= \sum_{i=1}^{m} \sum w_i V_{C_i}([\boldsymbol{a}, \boldsymbol{b}])
\end{aligned}$$

hence we see that since each component $C_i(u_1, u_2, \ldots, u_n)$ 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 $i \in \{1, 2, \ldots, m\}$. □

## Appendix D

**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

$$\begin{aligned}
q\left(\boldsymbol{\Upsilon}^{(t-1)}, \cdot\right) &= w_1 \mathcal{N}\left(\boldsymbol{\Upsilon}; \boldsymbol{\Upsilon}^{(t-1)}, \frac{(2.38)^2}{d} \mathbb{C}\mathrm{ov}\left(\left\{\boldsymbol{\Upsilon}^{(j)}\right\}_{0 \leq j \leq t-1}\right)\right) \\
&\quad + (1 - w_1) \mathcal{N}\left(\boldsymbol{\Upsilon}; \boldsymbol{\Upsilon}^{(t-1)}, \frac{(0.1)^2}{d} I_{d,d}\right),
\end{aligned} \tag{D.1}$$

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

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

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 $\widetilde{\Sigma} \in \mathrm{Sym}^+(d)$ in an adaptive MCMC proposal, restricted to the Riemann manifold of symmetric, positive definite $(d \times d)$ matrices, denoted by the space $\mathrm{Sym}^+(d)$.

**Remark D.1.** First, we note two properties of the marginal posterior $p\left(\widetilde{\Sigma} \,\middle|\, \left\{\widetilde{X}^P_{i,\mathrm{obs}}, \widetilde{X}^I_{i,\mathrm{obs}}\right\}_{0 \leq i \leq J}\right)$: 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

$$q\left(\widetilde{\Sigma}^{(t-1)}, \cdot\right) = w_1 \mathcal{IW}\left(\widetilde{\Sigma}; \Lambda_t^{\text{adap}}\left(\left\{\widetilde{\Sigma}^{(s)}\right\}_{0\le s\le t-1}\right), p\right) + (1-w_1)\,\mathcal{IW}\left(\widetilde{\Sigma}; \Lambda, p\right). \tag{D.3}$$

Here, the adaptive proposal mixture component is specified through fixing the degrees of freedom $p$ and then selecting $\Lambda_t^{\text{adap}}$ 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 (Eq. (D.4)), we note that we need to ensure that the sample average considered is restricted to the Riemann-manifold of positive definite matrices.

$$\Lambda_t^{\text{adap}}\left(\left\{\widetilde{\Sigma}^{(s)}\right\}_{0\le s\le t-1}\right) = \widehat{\widetilde{\Sigma}}^{(t-1)}\left(p - dim(\widetilde{\Sigma}) - 1\right). \tag{D.4}$$

This is satisfied through the choice of the estimator

$$\widehat{\widetilde{\Sigma}}^{(t-1)} = \frac{1}{t-1}\sum_{s=1}^{t-1}\widetilde{\Sigma}^{(s)}. \tag{D.5}$$

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 $\mathrm{Sym}^+(d)$.

## References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance Math. Econom. 44 (2), 182–198.
Andrieu, C., Thoms, J., 2008. A tutorial on adaptive MCMC. Stat. Comput. 18 (4), 343–373.
Dahms, R., 2008. A loss reserving method for incomplete claim data. Bull. Swiss Assoc. Actuar. 127–148.
Dahms, R., Merz, M., Wüthrich, M.V., 2009. Claims development result for combined claims incurred and claims paid data. Bull. Fr. d'Actuar. 9 (18), 5–39.
Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R., 2005. Actuarial Theory for Dependent Risks. Wiley Online Library.
Embrechts, P., 2009. Copulas: A personal view. J. Risk Insurance 76 (3), 639–650.
Gogol, D., 1993. Using expected loss ratios in reserving. Insurance Math. Econom. 12 (3), 297–299.
Gupta, A.K., Nagar, D.K., 2000. Matrix Variate Distributions. Vol. 104. Chapman & Hall/CRC.
Haario, H., Saksman, E., Tamminen, J., 2005. Componentwise adaptation for high dimensional MCMC. Comput. Statist. 20 (2), 265–273.
Happ, S., Wüthrich, M.V., 2013. Paid-incurred chain reserving method with dependence modeling. Astin Bull. 43 (1), 1–20.
Hertig, J., 1985. A statistical approach to the IBNR-reserves in marine insurance. Astin Bull. 15 (2), 171–183.
Hofert, Marius, Mächler, Martin, Mcneil, Alexander J., 2012. Likelihood inference for Archimedean copulas in high dimensions under known margins. J. Multivariate Anal. 110, 133–150.
Kannan, B., Lasscock, B., Mellen, C., Peters, G.W., Godsill, S., 2011. Bayesian cointegrated vector autoregression models incorporating alpha-stable noise for inter-day price movements via approximate Bayesian computation. Bayesian Anal. 6, 755–792.
Kannan, B., Lasscock, B., Peters, G.W., Mellen, C., 2010. Model selection and adaptive Markov chain Monte Carlo for Bayesian cointegrated var models. Bayesian Anal. 5, 465–492.
Kurowicka, D., Joe, H., 2010. Dependence Modeling: Vine Copula Handbook. World Scientific Publishing Co. Pvt. Ltd.
Mack, T., 1993. Distribution-free calculation of the standard error of chain ladder reserve estimates. Astin Bull. 23 (2), 213–225.
Merz, M., Wüthrich, M.V., 2010. Paid-incurred chain claims reserving method. Insurance Math. Econom. 46 (3), 568–579.
Merz, M., Wüthrich, M.V., 2013. Estimation of tail factors in the paid–incurred chain reserving method. Variance 7 (1), 61–73.
Min, A., Czado, C., 2010. Bayesian inference for multivariate copulas using pair-copula constructions. J. Financ. Econ. 8 (4), 511–546.
Nelsen, R.B., 2006. An Introduction to Copulas. Springer Verlag.
Patton, A.J., 2009. Copula-based models for financial time series. In: Handbook of Financial Time Series. pp. 767–785.
Peters, G.W., Wüthrich, M.V., Shevchenko, P.V., 2010. Chain ladder method: Bayesian bootstrap versus classical bootstrap. Insurance Math. Econom. 47 (1), 36–51.
Quarg, G., Mack, T., 2004. Munich chain ladder. Bl. DGVFM 26 (4), 597–630.
Roberts, G.O., Rosenthal, J.S., 2009. Examples of adaptive MCMC. J. Comput. Graph. Statist. 18 (2), 349–367.
Rosenthal, J.S., et al., 2008. Optimal proposal distributions and adaptive MCMC.
Stoica, P., Moses, R.L., 1997. Introduction to Spectral Analysis. Vol. 89. Prentice Hall, Upper Saddle River, NJ.
Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A., Tuerlinckx, F., 2011. Visualizing distributions of covariance matrices.
Zhang, Y., Dukic, V., 2013. Predicting multivariate insurance loss payments under the Bayesian copula framework. J. Risk Insurance 80 (4), 891–919.