WISHART-GAMMA MIXTURES FOR MULTIPERIL EXPERIENCE RATEMAKING, FREQUENCY-SEVERITY EXPERIENCE RATING AND MICRO-LOSS RESERVING

Denuit, M. and Y. LU

DISCUSSION PAPER | 2020 / 16

UCLouvain

ISBA | LIDAM Louvain Institute of Data Analysis and Modeling in economics and statistics

WISHART-GAMMA MIXTURES FOR MULTIPERIL EXPERIENCE RATEMAKING, FREQUENCY-SEVERITY EXPERIENCE RATING AND MICRO-LOSS RESERVING

Michel Denuit
Institute of Statistics, Biostatistics and Actuarial Science - ISBA
Louvain Institute of Data Analysis and Modeling - LIDAM
UCLouvain
Louvain-la-Neuve, Belgium
michel.denuit@uclouvain.be

Yang Lu
Department of Economics-CEPN
University of Paris 13
Villetaneuse, France
yang.lu@univ-paris13.fr

March 2, 2020

Abstract

This paper studies multivariate mixtures with Wishart-Gamma mixing distribution. After having recalled the definition and main properties of Wishart distributions for random symmetric positive definite matrices, it is shown how they can be used to extend Gamma distributions to the multivariate case, by considering the joint distribution of the diagonal terms. The resulting distribution, which we call Wishart-Gamma distribution, appears to be particularly useful to model correlated random effects in multivariate frequency, severity and duration models, leading to closed form likelihood function and posterior ratemaking formula. Three main applications are discussed to demonstrate the versatility of the Wishart-Gamma mixture models: (i) experience rating with several policies or guarantees per policyholder, (ii) experience rating taking into account the correlation between claim frequency and severity components, and (iii) dependence modeling between time-to-payment and amount of payment in micro-loss reserving when the ultimate payment is subject to censoring. Besides introducing the Wishart and Wishart-Gamma distributions, we are also among one of the first to employ the techniques such as fractional integral and symbolic calculation in the non-life actuarial literature.

Keywords: Mutlivariate Gamma distribution, Laplace transform, fractional calculus, symbolic calculation, Bayesian ratemaking, credibility, mixed Poisson distribution, frailty model.

1 Introduction and motivation

The Wishart distribution is meant for stochastic, symmetric, positive definite matrices. It appears to be the natural multivariate (in fact, matrix-variate) extension of the univariate Chi-square distribution. This distribution is named after John Wishart, who introduced it in 1928; see Wishart (1928). Some of the early theoretical advances concerning Wishart distributions have been published in one of the actuarial profession's oldest journals, the *Scandinavian Actuarial Journal*; see e.g. Sverdrup (1947) and Elfving (1947). It has been successfully applied in both finance (Gourieroux et al., 2009, Chiarella et al., 2014, Chiu and Wong, 2014) and life insurance (Cairns et al., 2006). In this paper, we show that Wishart distributions are also particularly useful in general insurance. Specifically, the diagonal elements of a matrix obeying the Wishart distribution define a tractable multivariate extension of the Gamma distribution that we call Wishart-Gamma law. This new extension of the univariate Gamma distribution can be efficiently combined with Poisson and Gamma parent distributions. Compared to other multivariate Gamma distributions recently proposed in the literature, those derived from Wishart distributions possess several important advantages. Firstly, the Laplace transform is available in closed form which eases inference and experience rating. Secondly, the parametrization by a scalar shape coefficient and a matrix scale coefficient appears to be quite flexible.

The multivariate Gamma distribution considered here has been proposed by Krishnamoorthy and Parthasarathy (1951) and further studied by Krishnaiah and Rao (1961) and Griffiths (1984). A special case of this distribution, with AR(1) covariance matrix, has been successfully applied to longitudinal count data modeling by Henderson and Shimakura (2003) in biostatistics and by Lu (2018) in insurance ratemaking. We are not aware of other applications of this multivariate Gamma distribution in the actuarial literature. Several competing multivariate extension of Gamma distributions have been considered in actuarial science:

(i) in a copula-based approach, Barges et al. (2009), Willmot and Woo (2012) and Shi and Valdez (2014) proposed to plug Gamma marginals into standard parametric copulas such as Gaussian, or Farlie-Gumble-Morgenstern. Let us also mention the related works by Frees and Wang (2005) and Yang and Shi (2019) who adopt a frequentist approach for multiperil insurance rating, that is, without introducing random effects. As argued by Frees and Wang (2005), the frequentist perspective avoids to make an assumption concerning the prior distribution of the latent variables. This may be seen as an advantage or disadvantage, depending on the situation. But model risk remains as copulas may be mis-specified, exactly as the structure function. The comparison between these two approaches is nevertheless beyond the scope of this paper.

(ii) Abdallah et al. (2016) proposed bivariate Gamma distributions in the Sarmanov family, in which the joint density is the product of the marginal Gamma densities, plus a correction term. The downside of this approach is that, first, it is difficult to extend this approach to higher dimensions. Secondly, even in dimension 2, the Sarmanov family of distributions is not sufficiently flexible since it can only accommodate for a limited range of correlation coefficients (as discussed e.g. by Ting Lee, 1996).

(iii) in a trivariate reduction-based approach, Alai et al. (2013) and Bermudez and Karlis (2017) proposed to sum three independent Gamma-distributed random variables Θ_1, Θ_2, and Θ_3 with the same scale parameter to obtain a correlated bivariate Gamma random vector $(\Theta_1+\Theta_2, \Theta_1+\Theta_3)$. This model, also called trivariate reduction, can be extended to higher dimensions, but a total of 2^n-1 independent Gamma-distributed random variables is required to generate n correlated Gamma variables. In comparison, the model proposed here is much more parsimonious. See however Furman and Landsman (2005) for a constrained multivariate specification.

The present paper demonstrates the strong potential of the Wishart distribution to solve problems encountered in general insurance with the help of several applications. Firstly, we propose a new, multivariate frequency regression model that appears to be effective for risk classification and experience rating. Specifically, the approach proposed here extends the seminal contribution by Dionne and Vanasse (1989) in the Poisson-Gamma setting for a single claim count to multiple, correlated claim counts. This appears to be particularly useful to simultaneously consider all the products subscribed by a given individual in a person-centric insurance approach, or all the products bought by individuals within a group (a household, for instance). Adopting the Wishart-Gamma structure function allows the actuary to derive closed-form premium updating formulas with a clear interpretation. This is in contrast with Pechon et al. (2018, 2019a,b)) who specified multivariate LogNormally-distributed random effects and resorted to numerical integration for calibration and premium updating.

The linear credibility approach, worked out in this setting e.g. by Pinquet (1998) and Englund et al. (2008) is easy to implement and more robust to model mis-specification (as discussed by Hong and Martin, 2019). However, it may fail to capture the nonlinearity of the pricing formula as demonstrated by Lu (2018). Recently, Pinquet (2019) investigated the conditions for the credibility coefficients to be positive. It turns out that these conditions are quite complicated and unless they are imposed ex ante, the updated premium potentially becomes negative.

This paper shows that using the Wishart-Gamma distribution to model multi-dimensional, unobservable random effects, preserves the mathematical tractability of the mixture model. This is in contrast with classical distributional assumptions such as multivariate LogNormal for instance. In other words, we generalize the well known Poisson-Gamma conjugacy to the multivariate case and show that the experience rating formulas remain explicit. Wishart-Gamma distributions appear to have two major advantages over competitors. Firstly, they are quite flexible even at higher dimensions, in the sense that they allow for various correlation patterns among the different random effect components. Secondly, model tractability does not come at the price of a complicated estimation procedure, as is usually the case for mixture models. For instance, Badescu et al. (2015) assume a mixture of Pascal distribution for the random effects (that is, Gamma distributions with integer degree of freedom) and proved that the likelihood function possesses a closed form formula. Lu (2019) established that the posterior distribution is also analytic, which facilitates experience rating. The downside of this approach, however, is that the approximation of the multivariate random effect distribution often requires a large number of components which complicates the calculation. It is argued that the approach proposed in this paper strikes a balance between flexibility and parsimony. Moreover, a simple composite likelihood estimation procedure for parameter estimation is available, which further increases the computational advantage of our approach. In the application, we propose a comparison of the premium updating formulas based on Wishart-Gamma structure function with those coming from linear credibility, when each policy has two guarantees. This comparison is particularly instructive, and shows that the mis-pricing of the credibility premium can sometimes be as high as ten percent for reasonable values of the parameters. Moreover, in another comparison between the closed form Bayesian premium formulas and their approximations using state-of-the-art approximation methods, it is shown that the approximation method might suffer both from computational cost and accuracy.

As a second application, a mixture model for correlated claim frequencies and severities is worked out. After Czado and Gschlossl (2007), several papers have relaxed the independence assumption between claim frequency and severity. Garrido et al. (2016) suggested to let the number of claims enter as a covariate in the severity model. Copulas have been used e.g. by Czado et al. (2012), Kramer et al. (2013), Shi et al. (2015), Lee and Shi (2019), Shi and Zhao (2019). Park et al. (2018) investigate possible reasons for the dependence between claim frequency and severity, like bonus hunger in motor insurance. See also Oh et al. (2019). Here, we follow Baumgartner et al. (2015), Jeong et al. (2017) and Lu (2019) work out a mixed model for dependent claim frequencies and severities using the multivariate Wishart-Gamma distribution as structure function. Again, our approach has the advantage of leading to closed form Bayesian ratemaking formula while at the same time being much easier to estimate than the model of Lu (2019).

As a third application, we consider loss reserving in general insurance based on micro-level data. Lopez (2019) suggested a copula model to describe the dependence between the amount of a claim and the time between its occurrence and its closure. This question is also relevant for multistate individual loss reserving models, as considered by Antonio et al. (2016) in a Markov setting and by Bettonville et al. (2020) in a semi-Markov setting. In this approach, the claim starts in the IBNR state and then evolves through a series of RBNS states until final settlement. The joint modeling of time to next payment and the corresponding amount, or link ratio is central to multistate micro-reserving models. A challenge here is that the payment is only observed in the absence of censoring, and for those censored claims, prediction of the final payment should be regularly updated until the payment happens, by taking into account the positive correlation between the time-to-payment and amount of payment. Again, we are not aware of any competing models that offer the same kind of tractability as the multivariate mixtures with Wishart-Gamma structure functions.

The remainder of the paper is structured as follows. Section 2 recalls the main properties of the Wishart distribution for random matrices and introduces the new Wishart-Gamma multivariate extension to the Gamma distribution. Relevant tools such as fractional calculus are also introduced in anticipation of the subsequent applications. The next sections are devoted to applications in nonlife insurance. Section 3 develops a multivariate frequency model with random effects obeying the Wishart-Gamma distribution. In Section 4, we propose a flexible modeling strategy for frequency and severity data, relaxing the commonly-made independence assumption. Section 5 is devoted to dependence modeling between the time-to-payment and the amount of payment in micro-reserving. It is also shown there that dependencies between lapse times for different products hold by the same individual can be modeled in a similar way. The final Section 6 discusses the results as well as possible extensions.

2 Wishart distribution

2.1 Multivariate extension of the Chi-square distribution

The Wishart distribution is a multivariate extension of the Gamma distribution defined over symmetric, positive-definite matrices with random elements. Specifically, consider independent random vectors $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ..., $\boldsymbol{Y}_\delta$ obeying the Normal distribution with zero mean vector and common variance-covariance matrix $\boldsymbol{\Sigma}$. The random matrix $\boldsymbol{S}$ defined as

$$\boldsymbol{S} = \boldsymbol{Y}_1\boldsymbol{Y}_1^\top + \boldsymbol{Y}_2\boldsymbol{Y}_2^\top + \cdots + \boldsymbol{Y}_\delta\boldsymbol{Y}_\delta^\top \tag{2.1}$$

is distributed according to the Wishart distribution with scale matrix $\boldsymbol{\Sigma}$ and number of degrees of freedom δ. Defined in this way, the Wishart distribution can be viewed as the matrix extension of the Chi-square distribution. It appears to be the sample distribution of variance-covariance matrix for multivariate Normal observations [see Anderson (1958)]. Finally, the joint Laplace transform of $\boldsymbol{S}$ is

$$\mathrm{E}[e^{-\mathrm{trace}(\boldsymbol{\Gamma}\boldsymbol{S})}] = \frac{1}{|\boldsymbol{I}_d + 2\boldsymbol{\Gamma}\boldsymbol{\Sigma}|^{1/2}}, \text{ for } \boldsymbol{\Gamma} \in \mathcal{S}_d^+,$$

where the determinant $|\boldsymbol{I}_d + 2\boldsymbol{\Gamma}\boldsymbol{\Sigma}|$ appearing in the denominator of (2.2) is always positive since it can be rewritten as

$$|\boldsymbol{I}_d + 2\boldsymbol{\Gamma}\boldsymbol{\Sigma}| = |\boldsymbol{I}_d + 2\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}| = |\boldsymbol{I}_d + 2\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{1/2}| > 0.$$

2.2 General definition

Exactly as Gamma distribution extends Chi-square with one degree of freedom to non-integer degrees of freedom, the general definition of Wishart law does not require δ to be an integer. Specifically, let $\mathcal{S}_d^+$ be the set of symmetric, positive semi-definite matrices $\boldsymbol{S}$ with dimension $d \times d$. The Wishart distribution with scale matrix $\boldsymbol{\Sigma} \in \mathcal{S}_d^+$ and δ degrees of freedom, denoted as Wishart$(\boldsymbol{\Sigma}, \delta)$, is generally defined in terms of its Laplace transform. Specifically,

$$\boldsymbol{S} \sim \text{Wishart}(\boldsymbol{\Sigma}, \delta) \text{ if } \mathrm{E}[e^{-\mathrm{trace}(\boldsymbol{\Gamma}\boldsymbol{S})}] = \frac{1}{|\boldsymbol{I}_d + 2\boldsymbol{\Gamma}\boldsymbol{\Sigma}|^{\delta/2}}, \text{ for } \boldsymbol{\Gamma} \in \mathcal{S}_d^+, \tag{2.2}$$

where in the last inequality we have used the fact that $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{1/2}$ is a symmetric, positive semi-definite matrix.

The characterization of the Wishart distribution in terms of its Laplace transform is the key to estimation and pricing in an appropriate multivariate random effect model. Considering Poisson parent distribution, all formulas only involve the Laplace transform that is available in closed-form in the Wishart case, as well as its derivatives. For other multivariate mixtures, such as those based on Gamma parent distribution, we will see that fractional derivatives of Laplace transforms enter actuarial calculations.

2.3 Properties

This section gathers some useful results about Wishart distributions. We refer to Andersen (1958) for a textbook treatment.

2.3.1 Link with univariate Gamma distribution

For any non-zero vector $\boldsymbol{\alpha} \in \mathbb{R}^d$, the random variable $\boldsymbol{\alpha}^\top \boldsymbol{S} \boldsymbol{\alpha}$ is non-negative since $\boldsymbol{S}$ is symmetric positive semi-definite. Let us show that it obeys the Gamma distribution. Recall that the probability density function of the Gamma distribution with mean a/b and variance a/b^2 is

$$f(x) = \frac{1}{\Gamma(a)} x^{a-1} b^a \exp(-bx) \text{ for } x \geq 0,$$

in the usual parametrization. Henceforth, we write $X \sim \text{Gamma}(a, b)$ to mean that the random variable X possesses this probability density function. The corresponding Laplace transform is

$$\mathrm{E}[e^{-uX}] = \left(1 + \frac{u}{b}\right)^{-a}.$$

Parameter a is referred to as the shape of the Gamma distribution whereas $1/b$ is known as its scale. The next property relates Wishart distributions for random matrices $\boldsymbol{S}$ to Gamma distributions for random variables $\boldsymbol{\alpha}^\top \boldsymbol{S} \boldsymbol{\alpha}$.

Property 2.1. *We have*

$$\boldsymbol{S} \sim \text{Wishart}(\boldsymbol{\Sigma}, \delta) \Rightarrow \boldsymbol{\alpha}^\top \boldsymbol{S} \boldsymbol{\alpha} \sim \text{Gamma}\left(\frac{\delta}{2}, \left(2\boldsymbol{\alpha}^\top \boldsymbol{\Sigma} \boldsymbol{\alpha}\right)^{-1}\right).$$

Proof. Let us derive the Laplace transform of the random variable $\boldsymbol{\alpha}^\top \boldsymbol{S} \boldsymbol{\alpha}$: formula (2.2) allows us to write

$$\begin{aligned} \mathrm{E}[e^{-u\boldsymbol{\alpha}^\top \boldsymbol{S} \boldsymbol{\alpha}}] &= \mathrm{E}[e^{-\mathrm{trace}(u\boldsymbol{\alpha}\boldsymbol{\alpha}^\top \boldsymbol{S})}] \\ &= \frac{1}{|\boldsymbol{I}_d + 2u\boldsymbol{\alpha}\boldsymbol{\alpha}^\top \boldsymbol{\Sigma}|^{\delta/2}} = \frac{1}{|\boldsymbol{I}_d + 2u\boldsymbol{\alpha}^\top \boldsymbol{\Sigma}\boldsymbol{\alpha}|^{\delta/2}} = \frac{1}{(1 + 2u\boldsymbol{\alpha}^\top \boldsymbol{\Sigma}\boldsymbol{\alpha})^{\delta/2}}, \end{aligned} \tag{2.3}$$

where in equation (2.3) we have used the property that $\boldsymbol{I}_m + \boldsymbol{A}\boldsymbol{B}$ and $\boldsymbol{I}_n + \boldsymbol{B}\boldsymbol{A}$ share the same determinant for any $(m \times n)$ matrix $\boldsymbol{A}$ and $(n \times m)$ matrix $\boldsymbol{B}$. This ends the proof since (2.3) corresponds to the Laplace transform of the Gamma distribution. □

Property 2.1 shows that every random variable of the form $\boldsymbol{\alpha}^\top \boldsymbol{S} \boldsymbol{\alpha}$ built from the random matrix $\boldsymbol{S}$ obeying the Wishart distribution follows the Gamma law with scale parameter $2\boldsymbol{\alpha}^\top \boldsymbol{\Sigma} \boldsymbol{\alpha}$ and shape parameter $\delta/2$. In particular, all diagonal elements S_{kk} of a random matrix $\boldsymbol{S}$ obeying the Wishart distribution are Gamma distributed. It suffices to choose

$\boldsymbol{\alpha}$ to be the $k-$th vector of the Euclidean canonical basis to deduce it from Property 2.1. Property 2.1 also shows that

$$\mathrm{E}[\boldsymbol{\alpha}^\top \boldsymbol{S}\boldsymbol{\alpha}] = \delta\boldsymbol{\alpha}^\top \boldsymbol{\Sigma}\boldsymbol{\alpha} \text{ and } \mathrm{Var}[\boldsymbol{\alpha}^\top \boldsymbol{S}\boldsymbol{\alpha}] = 2\delta(\boldsymbol{\alpha}^\top \boldsymbol{\Sigma}\boldsymbol{\alpha})^2.$$

2.3.2 Closure under marginalization

Another interesting property of the Wishart distribution is that it is closed under marginalization. As argued by Badescu et al. (2015) and Yang and Shi (2019), this property is highly desirable for general insurance. For instance, some clients might have several insurance policies (motor, home, etc), whereas others might only have one. Thus it is convenient, if the model used for all these clients are "compatible", regardless of the number of policies. In that respect, Wishart distributions are very convenient, as shown next.

Property 2.2. *If* $\boldsymbol{S} \sim \mathrm{Wishart}(\boldsymbol{\Sigma}, \delta)$ *then any of its sub-matrix* $(S_{i,j})_{i,j\in I}$*, where* I *is a subset of* $\{1, 2, ..., d\}$ *still follows Wishart distribution with dimension* $\#(I)$*.*

Proof. It suffices to take

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_1 & 0 \\ 0 & 0 \end{bmatrix},$$

where $\boldsymbol{\Gamma}_1$ is a $\#(I)$ dimensional symmetric positive semi-definite matrix and apply the Laplace transform formula (2.2). This ends the proof. □

2.3.3 Domain of δ

So far we have been silent on the domain of δ. According to Anderson (1958), if $\delta \in \{1, 2, ..., d-1\} \cup (d-1, \infty)$ then (2.2) defines a valid matrix distribution for any symmetric positive definite matrix parameter $\boldsymbol{\Sigma}$. When $\delta \in \{1, 2, ..., d-1\}$, the Wishart matrix is of reduced rank with probability 1, that is, $|\boldsymbol{S}| = 0$. These singular Wishart distributions have also been widely used in empirical works (see e.g. Uhlig, 1994). Moreover, although the whole matrix is "degenerate", other major properties, such as the Laplace transform formula (2.2) remain valid.

2.4 Multivariate Wishart-Gamma distribution

A multivariate extension of the Gamma distribution is obtained by considering the random vector with components corresponding to diagonal elements of a matrix obeying the Wishart distribution, which are all positive. Specifically, consider a random matrix $\boldsymbol{S} \sim \mathrm{Wishart}(\boldsymbol{\Sigma}, \delta)$ and define the random vector $\boldsymbol{X}$ as

$$X_k = S_{kk} \text{ for } k = 1, 2, \ldots, d.$$

This is henceforth denoted as $\boldsymbol{X} \sim \text{Wishart-Gamma}(\boldsymbol{\Sigma}, \delta)$. We know from Property 2.1 that each X_k is Gamma distributed (it suffices to choose $\boldsymbol{\alpha}$ to be the $k-$th vector of the Euclidean canonical basis). Specifically,

$$\boldsymbol{X} \sim \text{Wishart-Gamma}(\boldsymbol{\Sigma}, \delta) \Rightarrow X_k \sim \mathrm{Gamma}(\delta/2, 1/(2\sigma_{1,1})) \text{ for } k = 1, 2, \ldots, d.$$

The Laplace transform of $\boldsymbol{X}$ is available in closed form. Indeed, let us consider $\boldsymbol{\Gamma}$ to be diagonal, with $\boldsymbol{\Gamma} = \mathrm{diag}(u_1, u_2, ..., u_d)$. Equation (2.2) then gives

$$\mathrm{E}[e^{-u_1X_1-u_2X_2-\cdots-u_dX_d}] = \frac{1}{|\boldsymbol{I}_d + 2\mathrm{diag}(u_1, ..., u_d)\Sigma|^{\delta/2}}. \tag{2.4}$$

For instance, if $d = 2$, then we have:

$$\mathrm{E}[e^{-u_1X_1-u_2X_2}] = \frac{1}{\left[1 + 2u_1\sigma_{11} + 2u_2\sigma_{22} + 4u_1u_2(\sigma_{11}\sigma_{22} - \sigma_{12}^2)\right]^{\delta/2}}. \tag{2.5}$$

The domain of δ has been discussed in Section 2.3.3. The constraints on the parameters of the Wishart-Gamma distribution are sometimes even weaker. For instance, Griffiths (1984) established that

(i) If $d = 2$ then for any δ and any $\boldsymbol{\Sigma} \in \mathcal{S}_2^{++}$, equation (2.4) defines a valid matrix distribution.

(ii) If $d = 3$ and if the correlation matrix associated with $\boldsymbol{\Sigma}$ is an equi-correlation matrix with all the off-diagonal terms equal to some $\rho \in [0, 1]$, then the multivariate Gamma distribution exists for all values of δ.

Thus, we see that in the bivariate case, the Wishart-Gamma distribution is always defined even when the corresponding Wishart distribution may not, that is, when the Wishart-Gamma random vector cannot be embedded into a Wishart distributed matrix. For higher dimensions ($d \geq 3$), there can be some constrained specifications of the matrix parameter $\boldsymbol{\Sigma}$ for which the Wishart-Gamma distribution exists for all δ. Bernardoff (2006) provides some other sufficient (but unnecessary) conditions for the multivariate Wishart-Gamma distribution to be well defined.

Let us consider the particular cases of independence and perfect positive dependence, or comonotonicity.

Example 2.3. *If the matrix $\boldsymbol{\Sigma}$ is diagonal then equation* (2.4) *becomes*

$$\mathrm{E}[e^{-u_1X_1-u_2X_2-\cdots-u_dX_d}] = \frac{1}{(1 + 2u_1\sigma_{1,1})^{\delta/2}\cdots(1 + 2u_d\sigma_{d,d})^{\delta/2}}.$$

This shows that the components X_1, X_2, ..., X_d are independent, Gamma-distributed random variables with the same shape parameter. Thus the model includes the standard independent random effect model as a special case.

Example 2.4. *Assume now that $\boldsymbol{\Sigma}$ is degenerate with rank 1, that is, there exists a vector $\boldsymbol{\omega} \in \mathbb{R}^d$ such that*

$$\boldsymbol{\Sigma} = \boldsymbol{\omega}\boldsymbol{\omega}^\top.$$

For expository purpose, let us take $\boldsymbol{\omega} = (\sqrt{\sigma_{1,1}}, ..., \sqrt{\sigma_{d,d}})$ and investigate the resulting model.

If $\mathbf{\Gamma} = diag(u_1, u_2, ..., u_d)$ *then equation* (2.2) *gives*

$$\mathrm{E}[e^{-u_1X_1-u_2X_2-\cdots-u_dX_d}] = \frac{1}{(1+2\sigma_{1,1}u_1+\cdots+2\sigma_{d,d}u_d)^{\delta/2}}.$$

This corresponds to the joint Laplace transform of a degenerated distribution, that is,

$$X_2 = \frac{\sigma_{2,2}}{\sigma_{1,1}}X_1, \ldots, X_d = \frac{\sigma_{d,d}}{\sigma_{1,1}}X_1 \text{ almost surely,}$$

with X_1 *following the Gamma distribution. Thus the diagonal terms of a Wishart matrix admit the shared common factor model as a special case.*

As an immediate consequence of (2.5), we have

$$\boldsymbol{X} \sim \text{Wishart-Gamma}(\mathbf{\Sigma}, \delta) \Rightarrow \text{Corr}[X_i, X_j] = \frac{\sigma_{i,j}^2}{\sigma_{i,i}\sigma_{j,j}} \text{ for } i \leq j. \tag{2.6}$$

In other words, the correlation matrix of these diagonal elements is the elementwise square of the correlation matrix associated with the matrix $\mathbf{\Sigma}$. As a consequence, these correlations can take any values between 0 and 1, but cannot be negative. The Wishart-Gamma distribution thus only accomodates positive dependence relationships. Since positive correlation is supported by empirical evidence is many insurance applications, this is not a severe limitations. In the bivariate case, we will also see that negative dependence is easily accounted for, by inverting one of the two components.

Equation (2.6) illustrates the flexibility of the model since the pairwise correlation between (X_i, X_j) for different pairs (i, j) are captured by (the square of) different correlation parameters corresponding to the variance-covariance matrix $\mathbf{\Sigma}$. Moreover, this property is extremely convenient since it allows us to perform pairwise analysis and estimate the off-diagonal terms $\sigma_{i,j}$ separately for different couples (i, j) with $i \leq j$.

2.5 Wishart-Gamma distribution in random effect models

This subsection gives a preview of the type of models involved in the three case studied in Sections 3, 4, and 5. In particular we will see that under suitable specification, Wishart-Gamma based random effect models involve its joint Laplace transform (as well as its partial derivatives) only. These partial derivatives could be of integer, or fractional order. We will discuss how they can be computed in closed form in practice.

2.5.1 Models involving partial derivatives of integer orders only

Integer-ordered partial derivatives of the joint Laplace transform appears, for instance, in the context of multi-peril insurance ratemaking. Consider counting random variables $N_1, \ldots, N_d$ that are independent, given a positive random vector $(U_1, \ldots, U_d)$ and such that $N_j \sim \text{Poisson}(\lambda_j U_j)$ for each $j = 1, ..., d$. The joint probability mass function of $(N_1, \ldots, N_d)$ is then given by

$$\begin{aligned}
\mathrm{P}[N_1 = k_1, \dots, N_d = k_d] &= \frac{\lambda_1^{k_1} \dots \lambda_d^{k_d}}{k_1! \dots k_d!} \mathrm{E}\Big[U_1^{k_1} \dots U_d^{k_d} \exp\big(-\lambda_1 U_1 - \dots - \lambda_d U_d\big)\Big] \\
&= \frac{\lambda_1^{k_1} \dots \lambda_d^{k_d}}{k_1! \dots k_d!} (-1)^{k_1+\dots+k_d} \partial_1^{k_1} \dots \partial_d^{k_d} \mathrm{E}\Big[\exp\big(-\lambda_1 U_1 - \dots - \lambda_d U_d\big)\Big],
\end{aligned}$$

where $\partial_j^{k_j}$ denotes the k_j-th partial derivative with respect to the j-th argument, $j = 1, 2, \dots, d$. Hence, only the joint Laplace transform of the random vector $\boldsymbol{U}$ is needed to perform the calculation. If we assume that $\boldsymbol{U} \sim$ Wishart-Gamma$(\boldsymbol{\Sigma}, \delta)$ then this Laplace transform is given by (2.4). Note that the determinant $|\boldsymbol{I}_d + 2\mathrm{diag}(\lambda_{i,1,\bullet}, ..., \lambda_{i,d,\bullet})\boldsymbol{\Sigma}|$ is a polynomial function of the arguments $\lambda_{i,1,\bullet}$, ..., $\lambda_{i,d,\bullet}$, and hence the right-hand side of equation (3.1) is a fractional type function of $\lambda_{i,1,\bullet}$, ..., $\lambda_{i,d,\bullet}$ and the computation of the partial derivatives can be done rather straightforwardly by any symbolic computation package such as Mathematica. Moreover, in practical actuarial applications, most of these counts are either zero, or 1, thus the total order of the partial derivative is, most of the time, quite small and hence very fast to compute. Such computation has already been proposed in the literature by various authors, including Aalen (1987), Henderson and Shimakura (2003), and Lu (2018). For instance, for illustration purpose, the following lemma shows that such partial derivatives can be further simplified in the bivariate case.

Property 2.5. *For any a, b, $c \geq 0$, k_1, $k_2 \in \mathbb{N}$ and λ_1, $\lambda_2, \frac{\delta}{2}$ positive, we have*

$$\begin{aligned}
&\partial_1^{k_1} \partial_2^{k_2} \Big[\frac{1}{(1 + a\lambda_1 + b\lambda_2 + c\lambda_1\lambda_2)^{\frac{\delta}{2}}}\Big] \\
&= \frac{(-1)^{k_1+k_2} \Gamma(k_1+1)\Gamma(k_2+1)}{\Gamma(\frac{\delta}{2})} \\
&\quad \sum_{k=0}^{\min\{k_1,k_2\}} \frac{(-c)^k \Gamma(\frac{\delta}{2} + k_1 + k_2 - k)(a + c\lambda_2)^{k_1-k}(b + c\lambda_1)^{k_2-k}}{\Gamma(k_2 - k + 1)\Gamma(k_1 - k + 1)\Gamma(k+1)(1 + a\lambda_1 + b\lambda_2 + c\lambda_1\lambda_2)^{\frac{\delta}{2}+k_1+k_2-k}}.
\end{aligned}$$

Proof.

$$
\begin{aligned}
&\partial_1^{k_1}\partial_2^{k_2}\Big[\frac{1}{(1+a\lambda_1+b\lambda_2+c\lambda_1\lambda_2)^{\frac{\delta}{2}}}\Big]\\
&=(-1)^{k_2}\frac{\Gamma(\frac{\delta}{2}+k_2)}{\Gamma(\frac{\delta}{2})}\partial_1^{k_1}\Big[\frac{(b+c\lambda_1)^{k_2}}{(1+a\lambda_1+b\lambda_2+c\lambda_1\lambda_2)^{\frac{\delta}{2}+k_2}}\Big]\\
&=(-1)^{k_2}\frac{\Gamma(\frac{\delta}{2}+k_2)}{\Gamma(\frac{\delta}{2})}\sum_{k=0}^{k_1}\binom{k_1}{k}\frac{\Gamma(k_2+1)}{\Gamma(k_2-k+1)}\mathrm{I}[k_2\geq k]c^k(b+c\lambda_1)^{k_2-k}\\
&\qquad\qquad\frac{(-1)^{k_1-k}(a+c\lambda_2)^{k_1-k}\frac{\Gamma(\frac{\delta}{2}+k_1+k_2-k)}{\Gamma(\frac{\delta}{2}+k_2)}}{(1+a\lambda_1+b\lambda_2+c\lambda_1\lambda_2)^{\frac{\delta}{2}+k_1+k_2-k}}\\
&=\frac{(-1)^{k_1+k_2}\Gamma(k_1+1)\Gamma(k_2+1)}{\Gamma(\frac{\delta}{2})}\\
&\qquad\sum_{k=0}^{\min\{k_1,k_2\}}\frac{(-c)^k\Gamma(\frac{\delta}{2}+k_1+k_2-k)(a+c\lambda_2)^{k_2-k}(b+c\lambda_1)^{k_1-k}}{\Gamma(k_2-k+1)\Gamma(k_1-k+1)\Gamma(k+1)(1+a\lambda_1+b\lambda_2+c\lambda_1\lambda_2)^{\frac{\delta}{2}+k_1+k_2-k}},
\end{aligned}
$$

where we have used the identity $\binom{k_1}{k}=\frac{\Gamma(k_1+1)}{\Gamma(k+1)\Gamma(k_1-k+1)}$. This ends the proof. □

Note that the number of terms involved in the sum over k is equal to $1+\min\{k_1,k_2\}$. In particular, so long as one of k_1 or k_2 is zero, that is, when the policyholder has not declared any claim for either of the two products under consideration, the sum reduces to its first term:

$$
\partial_1^{k_1}\partial_2^{k_2}\Big[\frac{1}{(1+a\lambda_1+b\lambda_2+c\lambda_1\lambda_2)^{\frac{\delta}{2}}}\Big]=\frac{(-1)^{k_1+k_2}}{\Gamma(\frac{\delta}{2})}\frac{\Gamma(\frac{\delta}{2}+k_1+k_2)(a+c\lambda_2)^{k_1}(b+c\lambda_1)^{k_2}}{(1+a\lambda_1+b\lambda_2+c\lambda_1\lambda_2)^{\frac{\delta}{2}+k_1+k_3}}
$$

for all k_1 and k_2 such that $k_1k_2=0$.

2.5.2 Models involving fractional derivatives

In the above example, the orders of the partial derivatives are all nonnegative integers. However, in some other applications, fractional partial derivatives may naturally. For expository purpose let us focus on the special case where $d=2$ and consider the frequency-(total) severity joint modelling with Poisson and Gamma mixtures. Assume that (N,Y) are conditionally independent given (U_1,U_2), with Poisson(λ_1U_1) and Gamma$(\alpha N,\ \lambda_2U_2)$ distributions, respectively. Then the joint p.m.f./density of (N,Y) is, for any $n=0,1,..$ and $y\in\{0\}\cup\mathbb{R}_{>0}$:

$$
\mathrm{E}\left[\frac{(\lambda_1U_1)^ne^{-\lambda_1U_1}}{n!}\frac{(\lambda_2U_2)^{\alpha n}}{\Gamma(\alpha n)}y^{\alpha n}e^{-\lambda_2U_2ny}\right], \tag{2.7}
$$

with the convention that $0^{-1}=1$.

If α is integer, then the above expectation is nothing else but the partial derivatives or order (n,α), up to a sign; If, however, α is an arbitrary positive number, then the mathematical expression involved in the latter formula corresponds to the fractional derivative of the Laplace transform of $\boldsymbol{U}$. Let us now focus on this case in the reste of this subsection. Fractional calculs (integral or derivative) is well known in the mathematics literature (see e.g. Laue, 1980, for an early reference and Miller (1993) for a textbook treatment) but our paper provides one of its very first applications in actuarial science. More precisely, throughout the rest of this subsection we will prove the following property:

Property 2.6. *If (U_1, U_2) follows the bivariate Wishart-Gamma law with joint Laplace transform given by equation (2.5), then for any $s, t > 0$, and any exponent $\beta > -\delta/2$ we have:*

$$\mathrm{E}[e^{-sU_1 - tU_2} U_2^{\beta}] = \frac{\Gamma(\frac{\delta}{2} + \beta)}{\Gamma(\delta/2)} \frac{(b + cs)^{\beta}}{(1 + as + bt + cst)^{\frac{\delta}{2} + \beta}}, \tag{2.8}$$

where we have used the simplified notation $a = 2\sigma_{11}, b = 2\sigma_{22}, c = 4(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$.

To prove this property, let us first, for any univariate function f and $\beta > 0$, define its $\beta-$order Riemann-Liouville fractional integral $I^{\beta}(f)$ by:

$$I^{\beta}(f)(t) := \frac{1}{\Gamma(\beta)} \int_t^{\infty} (u - t)^{\beta - 1} f(u) \mathrm{d}u, \qquad \forall t \text{ within the domain of } f.$$

In particular, if $\beta = 1$, we recover the standard integral of function $f(\cdot)$. The following lemma provides the two equivalent expressions of the fractional integral of a Laplace transform.

Lemma 2.7 (The fractional integral of a Laplace transform)**.** *For any positive β and any positive random variable X, we have:*

$$\mathrm{E}[X^{-\beta} \exp(-tX)] = \frac{1}{\Gamma(\beta)} \int_t^{\infty} (u - t)^{\beta - 1} \mathrm{E}[e^{-uX}] \mathrm{d}u, \qquad \forall t > 0, \tag{2.9}$$

provided that both sides are well defined.

Proof of Lemma 2.7. By the definition of the gamma function, we have:

$$x^{-\beta} = \frac{1}{\Gamma(\beta)} \int_0^{\infty} s^{\beta - 1} e^{-sx} \mathrm{d}s, \qquad \forall x > 0.$$

Multiplying both sides by e^{-tx}, replacing the constant x by the random variable X, and taking expectation yields:

$$\mathrm{E}[X^{-\beta} \exp(-tX)] = \frac{1}{\Gamma(\beta)} \int_0^{\infty} s^{\beta - 1} \mathrm{E}[e^{-(t+s)X}] \mathrm{d}s$$

Then using the change of variable $u = t + s$, we get the right hand side of equation (2.9). $\square$

For general function f, the fractional integral does not necessarily have a closed form expression. Nevertheless, the Laplace transform of the gamma distribution is an exception with closed form fractional integral. More precisely, let us take $X \sim \text{Gamma}(\nu, \psi)$ for any positive ν, ψ. Then the fractional integral of the Laplace transform, i.e. $\mathrm{E}[X^{-\beta}\exp(-tX)]$ can be computed in two ways. On the one hand, straightforward algebra leads to:

$$\begin{aligned}\mathrm{E}[X^{-\beta}\exp(-tX)] &= \int_0^\infty x^{-\beta}\exp(-tx)\frac{x^{\nu-1}\exp(-\psi x)\psi^\nu}{\Gamma(\nu)}\mathrm{d}x \\ &= \frac{\psi^\nu}{\Gamma(\nu)}\int_0^\infty x^{\nu-\beta-1}\exp(-(\psi+t)x)\mathrm{d}x \\ &= \frac{\Gamma(\nu-\beta)}{\Gamma(\nu)}\frac{\psi^\nu}{(\psi+t)^{\nu-\beta}}, \qquad \forall t \geq 0, \end{aligned} \tag{2.10}$$

where the exponent β should be non larger than ν in order for both sides to be finite. On the other hand, by applying Lemma 2.7 we get:

$$\mathrm{E}[X^{-\beta}\exp(-tX)] = \frac{1}{\Gamma(\beta)}\int_t^\infty (u-t)^{\beta-1}\frac{\psi^\nu}{(t+\psi)^\nu}\mathrm{d}u. \tag{2.11}$$

By comparing equation (2.11) to (2.10) case we get a useful identity:

Lemma 2.8 (Fractional integral of a fractional type function)**.**

$$\int_t^\infty (u-t)^{\beta-1}\frac{1}{(u+\psi)^\nu}\mathrm{d}u = \frac{\Gamma(\nu-\beta)\Gamma(\beta)}{\Gamma(\nu)}\frac{1}{(\psi+u)^{\nu-\beta}}. \tag{2.12}$$

Let us now also establish an analog of Lemma 2.7 for bivariate Laplace transforms and its partial fractional integrals.

Lemma 2.9. *We have, for any positive β and any random vector (X, Y), we have:*

$$\mathrm{E}[e^{-sX-tY}Y^{-\beta}] = \frac{1}{\Gamma(\beta)}\int_t^\infty (v-t)^{\beta-1}\mathrm{E}[e^{-sX-vY}]\mathrm{d}v, \qquad \forall s,t>0, \tag{2.13}$$

provided that both sides are well defined.

The proof is similar to the previous lemma and is omitted. We now have all the necessary ingredients to prove Property 2.6.

Proof of Property 2.6. Let us remind that the joint Laplace transform of (U_1, U_2), by rewritting equation (2.5), is:

$$\mathrm{E}[e^{-sU_1-tU_2}] = \frac{1}{(1+as+bt+cst)^{\delta/2}} = \frac{1}{(b+cs)^{\frac{\delta}{2}}}\frac{1}{(t+\frac{1+as}{b+cs})^{\frac{\delta}{2}}}, \tag{2.14}$$

Thus we have, by applying Lemma 2.9:

$$\begin{aligned}
\mathrm{E}[e^{-sU_1-tU_2}U_2^{-\beta}] &= \frac{1}{\Gamma(\beta)}\int_t^\infty (v-t)^{\beta-1}\mathrm{E}[e^{-sU_1-vU_2}]\mathrm{d}v \\
&= \frac{1}{\Gamma(\beta)}\int_t^\infty (v-t)^{\beta-1}\frac{1}{(b+cs)^{\delta/2}}\frac{1}{(v+\frac{1+as}{b+cs})^{\frac{\delta}{2}}}\mathrm{d}v \\
&= \frac{1}{\Gamma(\beta)}\frac{1}{(b+cs)^{\delta}}\frac{\Gamma(\frac{\delta}{2}-\beta)\Gamma(\beta)}{\Gamma(\frac{\delta}{2})}\frac{1}{(t+\frac{1+as}{b+cs})^{\frac{\delta}{2}-\beta}} \qquad (2.15)\\
&= \frac{\Gamma(\frac{\delta}{2}-\beta)}{\Gamma(\frac{\delta}{2})}\frac{(c+\rho^2c^2s)^{-\beta}}{(1+cs+ct+\rho^2c^2st)^{\frac{\delta}{2}-\beta}}. \qquad (2.16)
\end{aligned}$$

where in equation (2.15) we have used the identity (2.12). Note that although in equation (2.15), we have implicitly used the constraint $\beta<\delta$ to make sure that $\Gamma(\beta)$ is well defined, this latter terms no longer appears in equation (2.16). On the other hand we can easily check that both the right hand side of equation (2.16), and $\mathrm{E}[e^{-sU_1-tU_2}U_2^{-\beta}]$, are analytic functions of β. Therefore, by analytic continuation, equation (2.16) still holds for negative β. Replacing β by $-\beta$ in this equation leads to Property 2.6, for any $\beta>-\delta/2$. □

Finally, once an expression of $\mathrm{E}[e^{-sU_1-tU_2}U_2^{\beta}]$ is found, differentiating it with respect to s yields the closed-form expression of $\mathrm{E}[U_1^nU_2^{-\beta}e^{-sU_1-tU_2}]$, This latter task again can be easily conducted using Mathematica. As a consequence, we have derived a closed form expression for the joint density defined in (2.7).

In both cases considered in this section, we see that formulas only involve the joint Laplace transform of the multivariate random effect $\boldsymbol{U}$ through its partial (regular or fractional) derivatives, which have been shown to have closed form when $\boldsymbol{U}$ obeys the Wishart-Gamma distribution. In the following we will consider three applications with mulviariate response variables, which could individually be of count type (with conditionally Poisson distribution), amount type (with conditionally gamma or Weibull distribution), or duration type (with conditionally time-changed exponential distribution as in the mixed proportional hazard model). Their combination leads to a large number of models that are tractable when appropriate properties of the Wishart-Gamma distribution are exploited. In the next three sections we review three quite different applications to illustrate the versatility of our approach.

3 Multiperil insurance rating

3.1 The Multivariate random effect model

Consider n policyholders, numbered $i=1,2,\dots,n$. Policyholder i may own up to d different insurance products, numbered $j=1,2,\dots,d$. They are followed over time t for m consecutive periods, numbered from 1 to m. The variables of interest are the numbers of claims $N_{i,j,t}$

recorded for policyholder i, product j and time period t. These counts obey the following assumptions:

Assumption 1 Given the positive random effects $\boldsymbol{U}_i = (U_{i,1}, U_{i,2}, ..., U_{i,d})$, the counts $N_{i,j,t}$ are independent and Poisson distributed with parameters $\lambda_{i,1,t}U_{i,1}, \cdots, \lambda_{i,d,t}U_{i,d}$, respectively, where $\lambda_{i,j,t}$ is the a priori claim frequency for product j, based on policyholder's i specific risk profile.

Assumption 2 The random effects $\boldsymbol{U}_1, ..., \boldsymbol{U}_n$ are independent and identically distributed across different individuals.

Assumption 3 Each $\boldsymbol{U}_i$ obeys the Wishart distribution with dimension d.

This multivariate model is a direct extension of the univariate framework proposed by Dionne and Vanasse (1989) to a multi-peril context.

3.2 Multiperil experience rating

In Subsection 2.5.1, we have already provided the closed form expression of the likelihood function of the above model. Let us show that the same model is also particularly convenient for Bayesian experience rating. Recall that the key property that the joint distribution of the random effects should satisfy in order to have closed form pricing formula is the availability of an analytical expression for the joint Laplace transform. This explains why Wishart-Gamma distributions are particularly appealing for ratemaking application.

In experience rating, the expected number of claims for next year is revised according to observed pas claims. For product j, this means that we are interested in (Bayesian) experience premium corresponding to the conditional expectation

$$\mathrm{E}[N_{i,j,m+1}|N_{i,k,t} \text{ for } k = 1, ..., d \text{ and } t = 1, ..., m].$$

Since random effects are constant over time, we know that this premium only depends on aggregate claim statistics, that is, on

$$N_{i,j,\bullet} = \sum_{t=1}^{m} N_{i,j,t} \text{ for } j = 1, 2, \ldots, d.$$

Specifically, the identity

$$\mathrm{E}[N_{i,j,m+1}|N_{i,k,t} \text{ for } k = 1, ..., d \text{ and } t = 1, ..., m] = \mathrm{E}[N_{i,j,m+1}|N_{i,k,\bullet} \text{ for } k = 1, ..., d]$$

holds true. Let us also define the corresponding aggregate expected claim frequencies as

$$\lambda_{i,j,\bullet} = \sum_{t=1}^{m} \lambda_{i,j,t} \text{ for } j = 1, 2, \ldots, d.$$

Given $\boldsymbol{U}_i$, we know that each $N_{i,j,\bullet}$ is Poisson distributed with mean $\lambda_{i,j,\bullet}U_{i,j}$ and that all these counts are mutually independent.

Then,

$$\begin{aligned}
&\mathrm{E}[N_{i,j,m+1}|N_{i,k,\bullet} \text{ for } k=1,...,d] \\
&= \lambda_{i,j,m+1}\mathrm{E}[U_{i,j}|N_{i,k,\bullet} \text{ for } k=1,...,d] \\
&= \lambda_{i,j,m+1}\frac{\mathrm{E}[U_{i,1}^{N_{i,1,\bullet}}\cdots U_{i,j}^{N_{i,j,\bullet}+1}\cdots U_{i,d}^{N_{i,d,\bullet}}e^{-\lambda_{i,1,\bullet}U_{i,1}-\cdots-\lambda_{i,d,\bullet}U_{i,d}}]}{\mathrm{E}[U_{i,1}^{N_{i,1,\bullet}}\cdots U_{i,j}^{N_{i,j,\bullet}}\cdots U_{i,d}^{N_{i,d,\bullet}}e^{-\lambda_{i,1,\bullet}U_{i,1}-\cdots-\lambda_{i,d,\bullet}U_{i,d}}]} \\
&= -\lambda_{i,j,m+1}\frac{\partial_1^{N_{i,1,\bullet}}\cdots\partial_j^{N_{i,j,\bullet}+1}\cdots\partial_d^{N_{i,d,\bullet}}\mathrm{E}[e^{-\lambda_{i,1,\bullet}U_{i,1}-\cdots-\lambda_{i,d,\bullet}U_{i,d}}]}{\partial_1^{N_{i,1,\bullet}}\cdots\partial_j^{N_{i,j,\bullet}}\cdots\partial_d^{N_{i,d,\bullet}}\mathrm{E}[e^{-\lambda_{i,1,\bullet}U_{i,1}-\cdots-\lambda_{i,d,\bullet}U_{i,d}}]},
\end{aligned} \tag{3.1}$$

where the joint Laplace transform is given by equation (2.4). Thus, the computation of the premium involves the partial derivatives of the joint Laplace transform, up to order $N_{i,1,\bullet}+\cdots+N_{i,d,\bullet}+1$, that is 1 plus the total number of claims reported during the m periods of observation, for all the d products under consideration.

3.3 Composite likelihood estimation for large d

Although the full likelihood function is available in closed form it involves the computation of a large dimensional determinant, if the number of products d is large, say larger than 3 or 4. The aim of this subsection is to propose a computationally simpler approach in this high-dimensional setting. This method is called composite likelihood estimation (see e.g. Varin et al., 2011, for a review). It has been widely applied to simplify the estimation, e.g. by Henderson and Shimakura (2003, Lu (2018), and Yang and Shi (2019). Roughly speaking, the idea of composite likelihood estimation is to replace the full likelihood function by simpler likelihood functions that correspond to "mis-specified" models depending on the same set of parameters. Considering the model satisfying Assumptions 1-3, the parameters are those involved in the a priori expected claim frequencies $\lambda_{i,j,t}$ as well as those of the Wishart-Gamma distribution, δ and $\mathbf{\Sigma}$.

The estimation approach is decomposed into two steps. First, we use the fact that each $U_{i,j}$ is Gamma distributed and $N_{i,j,t}$ is conditionally Poisson with parameter $\lambda_{i,j,t}U_{i,j}$. The marginal log-likelihood function is then given by

$$\ell_1 = \sum_{i=1}^{n}\sum_{j=1}^{d}\ell(N_{i,j,t},\ \ t=1,\ldots,m) \tag{3.2}$$

where $\ell(N_{i,j,t},\ \ t=1,\ldots,m)$ is the probability mass function of the random vector $(N_{i,j,1},\ldots,N_{i,j,m})$

which is Negative Binomial. Specifically,

$$
\begin{aligned}
\ell(N_{i,j,t},\ \ t=1,\dots,m) &\propto \mathrm{E}\Big[\prod_{t=1}^{m} e^{-\lambda_{i,j,t}U_{i,j}}(\lambda_{i,j,t}U_{i,j})^{N_{i,j,t}}\Big] \\
&= \Big(\prod_{t=1}^{m}\lambda_{i,j,t}^{N_{i,j,t}}\Big)(-1)^{N_{i,j,\bullet}}\frac{\mathrm{d}^{N_{i,j,\bullet}}}{\mathrm{d}\lambda_{i,j,\bullet}^{N_{i,j,\bullet}}}\Big(\frac{1}{(1+2\sigma_{j,j}\lambda_{i,j,\bullet})^{\delta/2}}\Big) \\
&= \Big(\prod_{t=1}^{m}\lambda_{i,j,t}^{N_{i,j,t}}\Big)\frac{\frac{\delta}{2}(\frac{\delta}{2}+1)(\frac{\delta}{2}+2)\cdots(\frac{\delta}{2}+N_{i,j,\bullet}-1)}{(1+2\sigma_{j,j}\lambda_{i,j,\bullet})^{\delta/2+N_{i,j,\bullet}}}
\end{aligned}
$$

where we have used the $\propto$ symbol to indicate that multiplicative constants that do not depend on the values of the parameters have been omitted. By maximizing function ℓ_1, we get an estimate of the parameters involved in the a priori expected claim frequencies $\lambda_{i,j,t}$ as well as of δ and the diagonal elements of $\mathbf{\Sigma}$.

It remains to estimate the off-diagonal elements of the matrix parameter $\mathbf{\Sigma}$. The key insight here is that for any $j_1, j_2 \in \{1, ..., d\}$ such that $j_1 < j_2$, coefficient σ_{j_1,j_2} measures the dependence between U_{i,j_1} and U_{i,j_2}, and hence also the dependence between $N_{i,j_1,t}$ and $N_{i,j_2,t}$. Thus by focusing jointly on two products j_1 and j_2, we should be able to identify σ_{j_1,j_2}. Specifically, we consider the pairwise log-likelihood function:

$$
\ell_{j_1,j_2} = \sum_{i=1}^{n} \ell\big((N_{i,j_1,t}, N_{i,j_2,t}),\ \ t=1,\dots,m\big) \tag{3.3}
$$

where $\ell((N_{i,j_1,t}, N_{i,j_2,t}),\ \ t=1,\dots,m)$ is the joint probability mass function of the random couples $(N_{i,j_1,t}, N_{i,j_2,t})$, $t=1,\dots,m$, that is given by

$$
\begin{aligned}
&\ell\big((N_{i,j_1,t}, N_{i,j_2,t}),\ \ t=1,\dots,m\big) \\
&\propto \Big(\prod_{t=1}^{m}\lambda_{i,j_1,t}^{N_{i,j_1,t}}\lambda_{i,j_2,t}^{N_{i,j_2,t}}\Big)\mathrm{E}[(U_{i,j_1})^{N_{i,j_1,\bullet}}(U_{i,j_2})^{N_{i,j_2,\bullet}}e^{-\lambda_{i,j_1,\bullet}U_{i,j_1}-\lambda_{i,j_2,\bullet}U_{i,j_2}}] \\
&= \Big(\prod_{t=1}^{m}\lambda_{i,j_1,t}^{N_{i,j_1,t}}\lambda_{i,j_2,t}^{N_{i,j_2,t}}\Big)(-1)^{N_{i,j_1,\bullet}+N_{i,j_2,\bullet}}\partial_{j_1}^{N_{i,j_1,\bullet}}\partial_{j_2}^{N_{i,j_2,\bullet}}\mathrm{E}[e^{-\lambda_{i,j_1,\bullet}U_{i,j_1}-\lambda_{i,j_2,\bullet}U_{i,j_2}}].
\end{aligned} \tag{3.4}
$$

We know from (2.4) that the joint Laplace transform $\mathrm{E}[e^{-\lambda_{i,j_1,\bullet}U_{i,j_1}-\lambda_{i,j_2,\bullet}U_{i,j_2}}]$ is given by

$$
\mathrm{E}[e^{-\lambda_{i,j_1,\bullet}U_{i,j_1}-\lambda_{i,j_2,\bullet}U_{i,j_2}}] = \frac{1}{\Big((1+2\lambda_{i,j_1,\bullet}\sigma_{j_1,j_1})(1+2\lambda_{i,j_2,\bullet}\sigma_{j_2,j_2}) - 4\lambda_{i,j_1,\bullet}\lambda_{i,j_2,\bullet}\sigma_{j_1,j_2}^2\Big)^{\delta/2}}
$$

and the partial derivatives in equation (3.4) are with respect to the arguments $\lambda_{i,j_1,\bullet}$ and $\lambda_{i,j_2,\bullet}$. Its expression is given by Property 2.5.

We can see that (3.3) only depends on one element of $\mathbf{\Sigma}$ through σ_{j_1,j_2}^2. Thus by maximizing the $\frac{p(p-1)}{2}$ single-argument, pairwise log-likelihood functions, we can get estimates of all terms σ_{j_1,j_2}. Rigorously speaking, the above pairwise composite likelihood only identifies σ_{j_1,j_2} up to a sign since the composite likelihood function ℓ_{j_1,j_2} depends on σ_{j_1,j_2} via its square. In the majority of insurance applications however, we focus on parameter $\mathbf{\Sigma}$ with positive off-diagonal terms only. The two important computational advantages of this pairwise approach is that first, these maximizations are with respect to one real argument and are therefore simple to perform; second, they can be done separately for different $j_1 < j_2$ as they depend on different elements σ_{j_1,j_2} of $\mathbf{\Sigma}$.

Finally, let us remark that although the above (pairwise) composite likelihood is very convenient, there is no theoretical assurance that the resulting estimated symmetric matrix parameter $\widehat{\mathbf{\Sigma}}$ is positive definite, even if by maximizing equation (3.2) we can constrain the parameter space of the diagonal elements of $\mathbf{\Sigma}$ to ensure that they are positive. There is thus a need for systematically checking whether $\widehat{\mathbf{\Sigma}}$ is positive definite, and this constitutes an excellent indication of the appropriateness of the Wishart-Gamma model for the data under consideration. For similar, recent pairwise analysis in high-dimensional problems, we refer the reader to Boloorforoosh et al. (2019) for a financial application and Gourieroux and Lu (2019) for an epidemiological application.

3.4 Bayesian Vs credibility premium for two products $(d = 2)$

Let us now consider the simplest example with two count variables $N_{i,1,t}$ and $N_{i,2,t}$ for each individual at each period t. We will see that in this case, the experience premium is available in closed form. To simplify the notations, we denote:

$$\begin{aligned} &N_{i,1,\bullet} = n_1, \qquad N_{i,2,\bullet} = n_2 \\ &2\sigma_{11} = a, \qquad 2\sigma_{22} = b, \qquad 4(\sigma_{11}\sigma_{22} - \sigma_{12}^2) = c \in [0, ab]. \end{aligned}$$

First, we compute the (n_1, n_2)-th, (n_1, n_2+1)-th, and (n_1+1, n_2)-th order partial derivatives of

$$\frac{1}{(1 + a\lambda_{i,1,\bullet} + b\lambda_{i,2,\bullet} + c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet})^{\delta/2}},$$

which are easily obtained by Property 2.5. For expository purpose, let us define $f(\cdot,\cdot)$ as

$$f(\lambda_{i,1,\bullet}, \lambda_{i,2,\bullet}) = \frac{1 + a\lambda_{i,1,\bullet} + b\lambda_{i,2,\bullet} + c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}}{(a + c\lambda_{i,2,\bullet})(b + c\lambda_{i,1,\bullet})}. \tag{3.5}$$

The experience premium (3.1) for, say, the first product can be expressed as

$$\begin{aligned} &\mathrm{E}[N_{i,1,m+1} | N_{i,1,\bullet} = n_1, N_{i,2,\bullet} = n_2] \\ &= \lambda_{i,1,m+1} \frac{\Gamma(n_1+2)}{\Gamma(n_1+1)} \frac{\sum_{k=0}^{\min\{n_1+1,n_2\}} \frac{(-c)^k \Gamma(\delta/2+n_1+n_2-k+1)(a+c\lambda_{i,2,\bullet})^{n_1+1-k}(b+c\lambda_{i,1,\bullet})^{n_2-k}}{\Gamma(n_2-k+1)\Gamma(n_1-k+2)\Gamma(k+1)(1+a\lambda_{i,1,\bullet}+b\lambda_{i,2,\bullet}+c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet})^{\delta/2+n_2+n_1-k+1}}}{\sum_{k=0}^{\min\{n_1,n_2\}} \frac{(-c)^k \Gamma(\delta/2+n_1+n_2-k)(a+c\lambda_{i,2,\bullet})^{n_1-k}(b+c\lambda_{i,1,\bullet})^{n_2-k}}{\Gamma(n_2-k+1)\Gamma(n_1-k+1)\Gamma(k+1)(1+a\lambda_{i,1,\bullet}+b\lambda_{i,2,\bullet}+c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet})^{\delta/2+n_2+n_1-k}}} \\ &= \lambda_{i,1,m+1} \frac{\Gamma(n_1+2)}{\Gamma(n_1+1)} \frac{(a + c\lambda_{i,2,\bullet})}{1 + a\lambda_{i,1,\bullet} + b\lambda_{i,2,\bullet} + c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}} \\ &\qquad \frac{\sum_{k=0}^{\min\{n_1+1,n_2\}} \frac{(-c)^k \Gamma(\delta/2+n_1+n_2-k+1)}{\Gamma(n_2-k+1)\Gamma(n_1-k+2)\Gamma(k+1)} f(\lambda_{i,1,\bullet}, \lambda_{i,2,\bullet})^k}{\sum_{k=0}^{\min\{n_1,n_2\}} \frac{(-c)^k \Gamma(\delta/2+n_1+n_2-k)}{\Gamma(n_2-k+1)\Gamma(n_1-k+1)\Gamma(k+1)} f(\lambda_{i,1,\bullet}, \lambda_{i,2,\bullet})^k}, \end{aligned} \tag{3.6}$$

As expected, the experience premium is a function of past individual aggregate expected claim frequencies $\lambda_{i,1,\bullet}$ and $\lambda_{i,2,\bullet}$, the future expected claim frequency $\lambda_{i,1,m+1}$, as well as of the total numbers of claims n_1 and n_2 reported in relation with the two products.

There are several special cases in which formula (3.6) can be further simplified. In the following we consider two of them: the model with independent random effects, as well as the experience premium for small values of n_1 and n_2 (as it is generally the case in insurance applications).

The independent model

Let us first consider the independent case, that is,

$$c = ab \Leftrightarrow \sigma_{12} = 0 \Leftrightarrow f(\lambda_{i,1,\bullet}, \lambda_{i,2,\bullet}) = \frac{1}{c}.$$

Instead of directly simplifying equation (3.6), notice that in this case the joint Laplace transform of $(U_{i,1}, U_{i,2})$ factors into:

$$\mathrm{E}[e^{-s_1 U_{i,1} - s_2 U_{i,2}}] = \frac{1}{(1+as_1)^{\delta/2}(1+bs_2)^{\delta/2}}, \tag{3.7}$$

therefore equation (3.1) becomes:

$$\mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = n_1, N_{i,2,\bullet} = n_2] = \lambda_{i,1,m+1}(\frac{\delta}{2} + d)\frac{a}{1 + a\lambda_{i,1,\bullet}}. \tag{3.8}$$

We thus recover the standard linear formula (in n_1) in the Poisson-Gamma conjugate case, as established by Dionne and Vanasse (1989).

Behavior of the experience premium around $(n_1, n_2) = (0, 0)$

Let us now consider the experience premium in the general case with dependence among the random effects of the two products, when the total numbers of claims n_1 and n_2 are small. First, if $n_2 = 0$, the summations in the above equation (3.6) reduce to their respective first terms, and we get:

$$\mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = n_1, N_{i,2,\bullet} = 0] = \lambda_{i,1,m+1}(\frac{\delta}{2} + n_1)\frac{(a + c\lambda_{i,2,\bullet})}{1 + a\lambda_{i,1,\bullet} + b\lambda_{i,2,\bullet} + c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}}$$

Thus, if there is no claim in relation to the second product (i.e. $N_{i,2,\bullet} = 0$), the experience premium is a linear function of n_1 with slope

$$\begin{aligned}&\mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = n_1 + 1, N_{i,2,\bullet} = 0] - \mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = n_1, N_{i,2,\bullet} = 0] \\ &= \lambda_{i,1,m+1}\frac{a + c\lambda_{i,2,\bullet}}{1 + a\lambda_{i,1,\bullet} + b\lambda_{i,2,\bullet} + c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}}\end{aligned} \tag{3.9}$$

Finally, by taking $n_1 = 0$ in equation (2.14), we get

$$\mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = N_{i,2,\bullet} = 0] = \lambda_{i,1,T+1}\frac{\delta}{2}\frac{a+\lambda_{i,2,\bullet}c}{1+a\lambda_{i,1,\bullet}+b\lambda_{i,2,\bullet}+c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}}. \tag{3.10}$$

Let us now compute the experience premium when no claim has been reported for the first product (i.e. $n_1 = 0$). In this case equation (3.6) becomes

$$\mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = 0, N_{i,2,\bullet} = n_2] = \lambda_{i,1,m+1}\frac{(a+c\lambda_{i,2,\bullet})}{1+a\lambda_{i,1,\bullet}+b\lambda_{i,2,\bullet}+c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}} \frac{\sum_{k=0}^{\min\{1,n_2\}}\frac{(-c)^k\Gamma(\frac{\delta}{2}+n_2-k+1)}{\Gamma(n_2-k+1)\Gamma(-k+2)\Gamma(k+1)}f(\lambda_{i,1,\bullet},\lambda_{i,2,\bullet})^k}{\frac{\Gamma(\frac{\delta}{2}+n_2)}{\Gamma(n_2+1)}}.$$

Then,

$$\begin{aligned}
&\mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = 0, N_{i,2,\bullet} = n_2]\\
&= \lambda_{i,1,m+1}\frac{(a+c\lambda_{i,2,\bullet})}{1+a\lambda_{i,1,\bullet}+b\lambda_{i,2,\bullet}+c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}}\frac{\frac{\Gamma(\frac{\delta}{2}+n_2+1)}{\Gamma(n_2+1)}-c\frac{\Gamma(\frac{\delta}{2}+n_2)}{\Gamma(n_2)}f(\lambda_{i,1,\bullet},\lambda_{i,2,\bullet})}{\frac{\Gamma(\frac{\delta}{2}+n_2)}{\Gamma(n_2+1)}}\\
&= \lambda_{i,1,m+1}\frac{(a+c\lambda_{i,2,\bullet})}{1+a\lambda_{i,1,\bullet}+b\lambda_{i,2,\bullet}+c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}}\Big((\frac{\delta}{2}+n_2)-cn_2f(\lambda_{i,1,\bullet},\lambda_{i,2,\bullet})\Big).
\end{aligned}$$

Thus the experience premium is again linear in n_2 when $n_1 = 0$, with slope

$$\begin{aligned}
&\mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = 0, N_{i,2,\bullet} = n_2+1] - \mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = 0, N_{i,2,\bullet} = n_2]\\
&= \lambda_{i,1,m+1}\frac{(a+c\lambda_{i,2,\bullet})}{1+a\lambda_{i,1,\bullet}+b\lambda_{i,2,\bullet}+c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}}\Big(1-cf(\lambda_{i,1,\bullet},\lambda_{i,2,\bullet})\Big) \qquad (3.11)\\
&= \lambda_{i,1,m+1}\frac{2\sigma_{1,2}^2}{\Big(1+2\sigma_{1,1}\lambda_{i,1,\bullet}+2\sigma_{2,2}\lambda_{i,2,\bullet}+4(\sigma_{1,1}\sigma_{2,2}-\sigma_{1,2}^2)\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}\Big)\Big(\sigma_{2,2}+2(\sigma_{1,1}\sigma_{2,2}-\sigma_{1,2}^2)\Big)}
\end{aligned}$$

For $n_1 = n_2 = 1$, equation (3.1) becomes

$$\begin{aligned}
&\mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = N_{i,2,\bullet} = 1]\\
&= \lambda_{i,1,m+1}\frac{\Gamma(3)}{\Gamma(2)}\frac{(a+c\lambda_{i,2,\bullet})}{1+a\lambda_{i,1,\bullet}+b\lambda_{i,2,\bullet}+c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}}\frac{\sum_{k=0}^{1}\frac{(-c)^k\Gamma(\frac{\delta}{2}+3-k)}{\Gamma(2-k)\Gamma(3-k)\Gamma(k+1)}f(\lambda_{i,1,\bullet},\lambda_{i,1,\bullet})^k}{\sum_{k=0}^{1}\frac{(-c)^k\Gamma(\frac{\delta}{2}+2-k)}{\Gamma(2-k)\Gamma(2-k)\Gamma(k+1)}f(\lambda_{i,1,\bullet},\lambda_{i,2,\bullet})^k}\\
&= 2\lambda_{i,1,m+1}\frac{(a+c\lambda_{i,2,\bullet})}{1+a\lambda_{i,1,\bullet}+b\lambda_{i,2,\bullet}+c\lambda_{i,1,\bullet}\lambda_{i,2,\bullet}}\frac{(\frac{\delta}{2}+1)(\frac{\delta}{2}+2)/2-c(\frac{\delta}{2}+1)f(\lambda_{i,1,\bullet},\lambda_{i,2,\bullet})}{(\frac{\delta}{2}+1)-cf(\lambda_{i,1,\bullet},\lambda_{i,2,\bullet})}.
\end{aligned}$$

From the above equation we can easily check that:

$$\mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = N_{i,2,\bullet} = 1] - \mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = 1, N_{i,2,\bullet} = 0] - \mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = 1, N_{i,2,\bullet} = 0]$$

$$+\mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = N_{i,2,\bullet} = 0] \neq 0,$$

that is, the Bayesian experience premium is a nonlinear function and the above term can be seen as the convexity adjustment in (n_1, n_2) at point $(n_1, n_2) = (0, 0)$.

To sum up, the experience premium depends on n_1 and n_2 in the following nonlinear way:

$$\mathrm{E}[N_{i,1,m+1}|N_{i,1,\bullet} = n_1, N_{i,2,\bullet} = n_2] = \omega_0 + \omega_1 n_1 + \omega_2 n_2 + g(n_1, n_2), \tag{3.12}$$

where ω_0, ω_1 and ω_2 do not depend on n_1 and n_2 and their values are given by equations (3.9), (3.10) and (3.11), respectively, whereas function $g(n_1, n_2)$ is the convexity adjustment with respect to $(0, 0)$, which is nonzero only if $n_1 n_2 \neq 0$. This result suggests that the weakness of the linear credibility approach is that it neglects the higher order, nonlinear adjustment terms. One could imagine that including, besides the linear terms in n_1 and n_2, some simple cross products such as $n_1 n_2$, might provide a lower forecasting error than the credibility premium (in terms of the mean square error). However, the downside of such an approach is that such a model does not guarantee the positivity of all the weight coefficients and thus can lead to potentially negative premium. See also Lu (2018) for similar discussions in a model with univariate but dynamic random effects.

Comparison with the linear credibility premium

Let us now compare the Bayesian posterior premium with the linear credibility premium. The latter is given by an affine function $\theta_0 + \theta_1 n_1 + \theta_2 n_2$ of past claims recorded for the two products under consideration. The coefficients θ_0, θ_1, and θ_2 are those of the linear regression model:

$$N_{i,1,m+1} = \theta_0 + \theta_1 N_{i,1,\bullet} + \theta_2 N_{i,2,\bullet} + \eta_{i,1,m+1} \tag{3.13}$$

with error term $\eta_{i,1,m+1}$ such that $\mathrm{E}[\eta_{i,1,m+1}|N_{i,1,\bullet}, N_{i,2,\bullet}] = 0$ or, stated equivalently,θ_0, θ_1, and θ_2 minimize the expected squared difference between the true expected claim frequency $\lambda_{i,1,m+1}U_{i,1}$ and its linear prediction $\theta_0 + \theta_1 N_{i,1,\bullet} + \theta_2 N_{i,2,\bullet}$. Then, it is straightforward to compute the theoretical values of these credibility coefficients, as shown next.

Property 3.1. *The credibility coefficients in equation* (3.13) *are given by:*

$$\theta_1 = \lambda_{i,1,m+1} \frac{a + \lambda_{i,2,\bullet}[ab - \frac{(ab-c)^2}{ab}]}{1 + a\lambda_{i,1,\bullet}\sigma_{2,2} + b\lambda_{i,2,\bullet} + \lambda_{i,1,\bullet}\lambda_{i,2,\bullet}[ab - \frac{(ab-c)^2}{ab}]}, \tag{3.14}$$

$$\theta_2 = \lambda_{i,1,m+1} \frac{\frac{ab-c}{b}}{1 + a\lambda_{i,1,\bullet} + b\lambda_{i,2,\bullet} + \lambda_{i,1,\bullet}\lambda_{i,2,\bullet}[ab - \frac{(ab-c)^2}{ab}]}, \tag{3.15}$$

$$\theta_0 = \lambda_{i,1,m+1} \frac{\delta}{2} \frac{a + c\lambda_{i,2,\bullet}}{1 + a\lambda_{i,1,\bullet} + b\lambda_{i,2,\bullet} + \lambda_{i,1,\bullet}\lambda_{i,2,\bullet}[ab - \frac{(ab-c)^2}{ab}]}. \tag{3.16}$$

Proof. By centering both sides of equation (3.13), multiplying by $N_{i,1,\bullet}$ or $N_{i,2,\bullet}$, and taking expectation, we get a linear system in (θ_1, θ_2):

$$\mathrm{Cov}[N_{i,1,m+1}, N_{i,1,\bullet}] = \theta_1 \mathrm{Var}[N_{i,1,\bullet}] + \theta_2 \mathrm{Cov}[N_{i,1,\bullet}, N_{i,2,\bullet}], \tag{3.17}$$

$$\mathrm{Cov}[N_{1,1,m+1}, N_{i,2,\bullet}] = \theta_1 \mathrm{Cov}[N_{i,1,\bullet}, N_{i,2,\bullet}] + \theta_2 \mathrm{Var}[N_{i,2,\bullet}]. \tag{3.18}$$

The (co-)variances involved in the above two equations are easily computed using the iterative expectation formula and the fact that given the random effects $U_{i,1}$ and $U_{i,2}$, counts $N_{i,1,m+1}$, $N_{i,1,\bullet}$, $N_{i,2,\bullet}$ are conditionally independent and Poisson distributed with parameter $\lambda_{i,1,m+1}U_{i,1}$, $\lambda_{i,1,\bullet}U_{i,1}$ and $\lambda_{i,2,\bullet}U_{i,2}$, respectively. This gives

$$\begin{aligned}
\mathrm{Cov}[N_{i,1,m+1}, N_{i,1,\bullet}] &= \lambda_{i,1,m+1}\lambda_{i,1,\bullet}2\sigma_{1,1}^2\delta \\
\mathrm{Cov}[N_{i,1,m+1}, N_{i,2,\bullet}] &= \lambda_{1,1,m+1}\lambda_{i,2,\bullet}2\sigma_{1,2}^2\delta \\
\mathrm{Var}[N_{i,1,\bullet}] &= \lambda_{i,1,\bullet}^2 2\sigma_{1,1}^2\delta + \lambda_{i,1,\bullet}\sigma_{1,1}\delta \\
\mathrm{Var}[N_{i,2,\bullet}] &= \lambda_{i,2,\bullet}^2 2\sigma_{2,2}^2\delta + \lambda_{i,2,\bullet}\sigma_{2,2}\delta \\
\mathrm{Cov}[N_{i,1,\bullet}, N_{i,2,\bullet}] &= \lambda_{i,1,\bullet}\lambda_{i,2,\bullet}2\sigma_{1,2}^2\delta.
\end{aligned}$$

By solving the system (3.17)-(3.18) we get the announced expressions (3.14)-(3.15) for the coefficients θ_1 and θ_2. Finally, by taking expectation in equation (3.13) we get the announced expression (3.16) for θ_0. This ends the proof. □

The values of θ_1 and θ_2 given in (3.14) and (3.15), respectively, are the slope of the linear credibility premium. Compared with the Bayesian experience premium, given by equations (3.9) and (3.11), respectively, we see that slopes are generally different. However, we would have

$$d_1 = \mathrm{E}[N_{i,1,m+1}|n_1 + 1, n_2 = 0] - \mathrm{E}[N_{i,1,m+1}|n_1, n_2 = 0]$$

and

$$d_2 = \mathrm{E}[N_{i,1,m+1}|n_1 = 0, n_2 + 1] - \mathrm{E}[N_{i,1,m+1}|n_1 = 0, n_2],$$

by replacing c by ab in equations (3.14) and (3.15), that is when the two random effects U_1, U_2 are independent. Similarly, the linear credibility premium for an individual with no previous accidents ($n_1 = n_2 = 0$), which is d_0 given in (3.16), is not equal to the Bayesian experience premium for the same individual, which is given by equation (3.10). Indeed the right-hand sides of equations (3.16) and (3.10) share the same, positive numerator; However their denominators differ. Indeed, since

$$ab - \frac{(ab-c)^2}{ab} \geq c$$

the Bayesian experience premium is higher than the linear credibility premium for an individual with $(n_1, n_2) = (0, 0)$.

As a summary, the main findings regarding the comparison between the Bayesian experience premium and its linear credibility approximation is that

(i) their values at $(0, 0)$ are not equal.

(ii) the linear credibility premium can be viewed as a linear approximation of the exact, Bayesian experience premium. It is a linear function in both arguments, whereas the Bayesian experience premium is nonlinear. Nevertheless, the experience premium is locally linear at $(0, 0)$, but the slopes are not equal to those of the linear credibility premium.

As an illustration, in the following, we provide a 3D comparison between the Bayesian and credibility ratemaking function in n_1, n_2. For diffent values of n_1, n_2, we compute the relative percentage of underpricing by the credibility premium, defined by:

$$\text{Percentage of underpricing} = \frac{\text{Bayesian premium} - \text{credibility premium}}{\text{Bayesian premium}}$$

The parameter values of the model we use are:

$$\lambda_{i,1,\bullet} = 0.17, \lambda_{2,1,\bullet} = 0.2, a = b = 1, c = 0.6, \delta = 1.$$

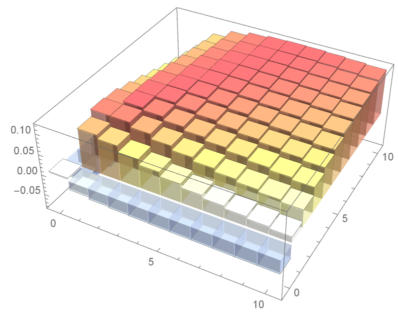

Figure 1: Relative percentage of underpricing by the credibility method compared to the exact Bayesian premium, for $\lambda_{i,1,\bullet} = 0.17, \lambda_{2,1,\bullet} = 2, a = b = 1, c = 0.6$, $\delta = 1$. The range of n_1, n_2 are set to $\{0, 1, ..., 10\}$.

As we can see, as expected, at $(0, 0)$, the credibility formula induces a (slight) underpricing. However it induces an overpricing, if either of the two arguments increases while the other remain zero. On the other hand, when both n_1 and n_2 are positive, the underpricing again becomes significant, exceeding 10% for $(n_1 = n_2) = (1, 1)$. Note also that these percentage errors depend also on the values of the parameters. Below we conduct the same benchmarking between credibility and Bayesian expectation for three alternative sets of parameters and understand the sensitivility of the pricing error with respect to different parameters. We are particularly interested in two paramters. The first one is $c = 4(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$, which, for given values of $a = 2\sigma_{11}$ and $b = 2\sigma_{22}$, characterizes the correlation between $U_{i,1}$ and $U_{i,2}$. The second ones are $\lambda_{i,1,\bullet}$ and $\lambda_{i,2,\bullet}$, which increase when the number of observed periods increases. The major conclusions of these comparisons are provided in bold letters in the captions below.

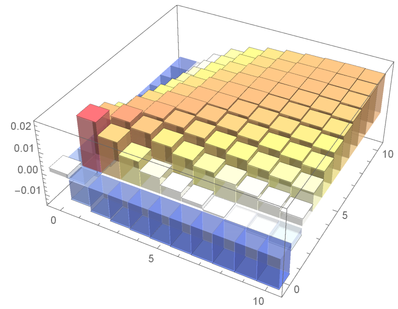

Figure 2: Relative percentage of underpricing by the credibility method compared to the exact Bayesian premium, with $c = 0.1$ and the other parameters are the same except as in Figure 1. As we can see the error is globally smaller than in Figure 1, which is expected, since in the limiting case where $c = 0$, the error is zero since in this independent case the credibility premium is exact.

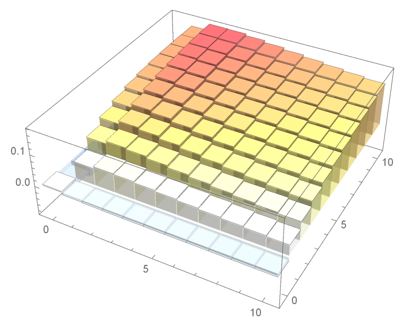

Figure 3: Relative percentage of underpricing by the credibility method compared to the exact Bayesian premium, with $c = 0.9$ and the other parameters are the same except as in Figure 1. This corresponds to the case where the two random effects are highly correlated (the limiting case where $c = 1$ corresponds to the comonotonicity). We can see that the errors are generically larger than in Figure 1. Thus by comparing Figures 1, 2, and 3, we conclude that **the stronger the dependence, the larger the pricing error**.

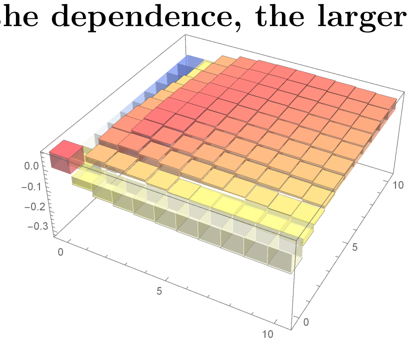

Figure 4: Relative percentage of underpricing by the credibility method compared to the exact Bayesian premium, with $\lambda_{i,1,\bullet} = 1.7, \lambda_{2,1,\bullet} = 2$, that is ten times larger than those chosen for Figure 1. The other parameters are the same as in Figure 1. We can see that the pricing error is generically larger than in Figure 1. In other words, **the longer the observation period, the larger the pricing error**.

3.5 Numerical illustration

In a series of papers, Pechon et al. (2018, 2019a,b) have studied claim frequencies on all motor and home products owned by policyholders from the same household. This has been performed by fitting a multivariate Poisson-LogNormal mixture to the $d = 7$ corresponding counts: number of claims reported by the household in home insurance ($j = 1$), number of claims reported by husband that triggered *only* motor third-party liability insurance ($j = 2$), number of claims reported by husband that triggered *only* material damage insurance ($j = 3$), number of claims reported by husband that triggered *both* motor third-party liability insurance and material damage insurance ($j = 4$), number of claims reported by wife that triggered *only* motor third-party liability insurance ($j = 5$), number of claims reported by wife that triggered *only* material damage insurance ($j = 6$), and number of claims reported by wife that triggered *both* motor third-party liability insurance and material damage insurance ($j = 7$). Here, the index i refers to the household.

In this section, we work with the following correlation structure, taken from Pechon et al. (2019b). The corresponding random effects have unit mean, variances

$$\begin{aligned}
\mathrm{Var}[U_{i,1}] &= 0.3655 \\
\mathrm{Var}[U_{i,2}] &= \mathrm{Var}[U_{i,5}] = 0.7406 \\
\mathrm{Var}[U_{i,3}] &= \mathrm{Var}[U_{i,6}] = 0.4387 \\
\mathrm{Var}[U_{i,4}] &= \mathrm{Var}[U_{i,7}] = 0.3914
\end{aligned}$$

and correlations

	$U_{i,1}$	$U_{i,2}$	$U_{i,3}$	$U_{i,4}$	$U_{i,5}$	$U_{i,6}$	$U_{i,7}$
$U_{i,1}$	1.0000	0.2457	0.3193	0.3380	0.2457	0.3193	0.3380
$U_{i,2}$	0.2457	1.0000	0.4165	0.4410	0.3206	0.4165	0.4410
$U_{i,3}$	0.3193	0.4165	1.0000	0.5730	0.4165	0.5412	0.5730
$U_{i,4}$	0.3380	0.4410	0.5730	1.0000	0.4410	0.5730	0.6066
$U_{i,5}$	0.2457	0.3206	0.4165	0.4410	1.0000	0.4165	0.4410
$U_{i,6}$	0.3193	0.4165	0.5412	0.5730	0.4165	1.0000	0.5730
$U_{i,7}$	0.3380	0.4410	0.5730	0.6066	0.4410	0.5730	1.0000

One particularly interesting Wishart-Gamma model is its special case where $\delta = 1$. Indeed, in this case we have said that $(U_{i,1}, ..., U_{i,7})$ has the representation of the componentwise square of a zero mean Gaussian vector. Moreover, by Property 2.1 and equation (2.6), the covariance matrix of this Gaussian vector is given by the componentwise square root of the covariance matrix of $(U_{i,1}, ..., U_{i,7})$, divided by $\sqrt{2}$. Therefore, under this model, the joint Laplace transform and its derivatives in equation (3.1) can be evaluated in two ways. Either we can compute them in closed form through Mathematica, or alternatively we can use the Gauss-Hermite quadrature method as advocated by Pechon (2018a). Thus this model provides an excellent way of demonstrating the accuracy and computational advantage of our approach compared to the state-of-the-art approximation approach. Below we imagine a hypothetical policyholder observed over one period only, with a prior scores $\lambda_{i,j,t=1} = 0.1$ for each $j = 1, ...7$, with no observed claim: $N_{i,j,t=1} = 0$ for each j. We are interested in the posterior expectation of $N_{i,j=1,t=2}$, that is the expected number of claims for home insurance. We express this posterior expectation as a multiple of $\lambda_{i,j,t=2}$, that is the "Bayesian credibility coefficient", which is smaller than one to reflect the fact that this policyholder will get a deduction at the second period. We compute the Bayesian credibility coefficient using Mathematica, and then using the same R package for Gauss-Hermite quadratures as in Pechon (2018a), for different values of m, that is the number of quadrature points *per dimension*. In other words, since we deal with a 7-dimensional Gaussian distribution, the total number of quadrature points is m^7. The model of Pechon (2018a) involves a bivariate Gaussian distribution and there, they recommended to use $m \geq 15$ to achieve an excellent approximation accuracy. In our case since the dimension is significantly higher, we were only able to compute the Gauss-Hermite quadratures for m up to 10 if we were to keep the computational time reasonable. Their approximation error and CPU times are reported below:

	Value of the premium	CPU time	Percentage error
Exact value obtained Mathematica	0.6963	< 0.01s	0, by definition
Gauss-Hermite $m = 10$	0.6964	83s	0.01 %
Gauss-Hermite $m = 9$	0.6959	33s	0.1%
Gauss-Hermite $m = 8$	0.6973	12s	0.5 %
Gauss-Hermite $m = 7$	0.6934	4s	0.4 %
Gauss-Hermite $m = 6$	0.7037	1.4s	1 %

Table 3.1: Comparison between the exact method and the Gauss-Hermite quadrature method for the computation of the credibility coefficient for an individual without claim in the first period.

We can see that the approximation error becomes negligible, only when m is taken to be at least, say, 9, which leads to a high computational cost. This is expected since the approximation scheme has a complexity of m^7 which increases very quickly in m.[1] As a summary, the Wishart-Gamma specification, along with its closed form formulas, renders the pricing (and estimation) much quicker than when alternative models are used and approximation methods have to be employed.

4 Correlated claim frequencies and severities

4.1 The model

Assume the number of claims $N_{i,t}$ and the total amount of claims $Y_{i,t}$ are available for each policyholder i and time period t in relation to a product of interest. It is often assumed that $N_{i,t}$ and the individual claim severities $C_{i,k,t}$ aggregated into $Y_{i,t}$ are mutually independent.

[1] We are aware that there are parallelisation possibilities and more powerful computing softwares such as C++ [see Pechon (2019b)] and their implementation would make the approximation method less cumbersome. However we have not pursued this route given the significant investment and expertise required. As a consequence we have not optimized our Mathematica program either, in order to ensure a "fair" horse race between the two approaches.

However, this assumption has been questioned based on empirical evidence suggesting that some correlation may be present. Even if the reasons underlying this dependence are still in debate, it should be taken into account in pricing because the pure premium cannot be computed as the product of the expected claim frequency times the expected claim severity. Moreover, throughout this section, we assume that:

Assumption 1bis Given the random effects $(U_{i,1}, U_{i,2})$, the claim counts $N_{i,t}$ are independent and Poisson distributed with parameter $\lambda_{i,1,t}U_{i,1}$.

Assumption 2bis The total claim amount $Y_{i,t}$ admits the representation $Y_{i,t} = \sum_{k=1}^{N_{i,t}} C_{i,k,t}$ where $C_{i,1,t}$, $C_{i,2,t}$, $\ldots$ are the individual claim severities at period t for individual i. Given the random effects $(U_{i,1}, U_{i,2})$, claim sizes $C_{i,k,t}$ are independent, gamma distributed with rate parameter $\lambda_{i,2,t}U_{i,2}$ and shape parameter α.

Assumption 3bis Given the random effects $(U_{i,1}, U_{i,2})$, all random variables $N_{i,t}$, $C_{i,1,t}$, $C_{i,2,t}$, $\ldots$, $t = 1, 2, \ldots$, are mutually independent.

Assumption 4bis The random effects $\boldsymbol{U}_i = (U_{i,1}, U_{i,2})$ are independent and distributed according to the Wishart-Gamma law.

Thus, according to Assumption 2bis, $Y_{i,t}$ is conditionally Gamma$(\alpha N_{i,t}, \lambda_{i,2,t}U_{i,2})$ given $U_{i,2}$ and $N_{i,t}$, or conditionally Tweedie given $U_{i,2}$ only. Let us also remind the standard gamma-gamma conjugacy. That is, since $U_{i,2}$ follows Gamma$(\delta/2, \delta)$, the marginal distribution of each individual claim at date t has the density:

$$
\begin{aligned}
\ell(y|\lambda_{i,2,t}) &:= \mathrm{E}\Big[\frac{(\lambda_{i,2,t}U_{2,i})^\alpha}{\Gamma(\alpha)}\exp(-\lambda_{i,2,t}U_{2,i}y)y^{\alpha-1}\Big] \\
&= \frac{\lambda_{i,2,t}^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \int e^{-\lambda_{i,2,t}u_2 y} u_2^\alpha \frac{u_2^{\frac{\delta}{2}-1} e^{-u_2/b}}{\Gamma(\frac{\delta}{2}) b^{\frac{\delta}{2}}} \mathrm{d}u_2 \\
&= \frac{\lambda_{i,2,t}^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+\frac{\delta}{2})}{\Gamma(\frac{\delta}{2}) b^{\frac{\delta}{2}}} \frac{y^{\alpha-1}}{(\lambda_{i,2,t}y + \frac{1}{b})^{\alpha+\frac{\delta}{2}}}, \qquad \forall y > 0,
\end{aligned}
$$

which is the p.d.f. of a generalized beta distribution of the second kind (GB2) and is widely used in the (a prior) ratemaking literature [see Cummins et al. (1990)]. It is well known that this distribution has a finite mean if and only if $\frac{\delta}{2} > 1$. Thus throughout this section we will assume that $\delta > 2$.

The conditional expected frequency and total cost are respectively given by

$$
\mathrm{E}[N_{i,t}|U_{i,1}, U_{i,2}] = \lambda_{i,1,t}U_{i,1} \text{ and } \mathrm{E}[Y_{i,t}|U_{i,1}, U_{i,2}, N_{i,t}] = \alpha \frac{N_{i,t}}{\lambda_{i,2,t}U_{2,i}}.
$$

The pure premium is

$$\begin{aligned}
\mathrm{E}[Y_{i,t}] &= \mathrm{E}\left[\mathrm{E}\left[\sum_{k=1}^{N_{i,t}} C_{i,k,t} \middle| U_{i,1}, U_{i,2}\right]\right] \\
&= \alpha \frac{\lambda_{i,1,t}}{\lambda_{i,2,t}} \mathrm{E}\left[\frac{U_{i,1}}{U_{i,2}}\right].
\end{aligned}$$

Depending on the correlation structure within the pair $(U_{i,1}, U_{i,2})$, the product of the expected claim number and of the expected cost per claim may under- or over-estimates the pure premium (both a priori and a posteriori). Under Assumptions 1bis-4bis, the expected frequency and expected average costs $Y_{i,t}/N_{i,t}$ are negatively correlated, if $U_{i,1}$ and $U_{i,2}$ are positively correlated, as it is the case when $\boldsymbol{U}_i$ follows the Wishart-Gamma distribution. Let us now compute the term $\mathrm{E}[\frac{U_{i,1}}{U_{i,2}}]$, which is the value of $\mathrm{E}[\frac{U_{i,1}}{U_{i,2}} e^{-sU_{i,1}-tU_{i,2}}]$ when $s = t = 0$. It can be computed as follows. First, since $-1 > -\frac{\delta}{2}$, we can apply Property 2.6 and get:

$$\mathrm{E}\Big[\frac{1}{U_2} e^{-sU_{i,1}-tU_{i,2}}\Big] = \frac{\Gamma(\frac{\delta}{2}-1)}{\Gamma(\frac{\delta}{2})} \frac{(b+cs)^{-1}}{(1+as+bt+cst)^{\frac{\delta}{2}-1}} = \frac{1}{\frac{\delta}{2}-1} \frac{(b+cs)^{-1}}{(1+as+bt+cst)^{\frac{\delta}{2}-1}}. \quad (4.1)$$

Then taking derivative with respect to s yields:

$$\mathrm{E}\Big[\frac{U_{i,1}}{U_{i,2}} e^{-sU_{i,1}-tU_{i,2}}\Big] = \frac{1}{\frac{\delta}{2}-1} \frac{\frac{c}{(b+cs)^2}(1+as+bt+cst) + (\frac{\delta}{2}-1)\frac{a+ct}{b+cs}}{(1+as+bt+cst)^{\frac{\delta}{2}}}, \quad (4.2)$$

Hence by taking $s = t = 0$ in the above equation we deduce:

$$\mathrm{E}\Big[\frac{U_{i,1}}{U_{i,2}}\Big] = \frac{1}{\frac{\delta}{2}-1}\Big[\frac{c}{b^2} + (\frac{\delta}{2}-1)\frac{a}{b}\Big]. \quad (4.3)$$

It is easily checked that $\mathrm{E}[U_{i,1}] = \frac{\delta}{2}a$ and $\mathrm{E}[\frac{1}{U_{i,2}}] = \frac{1}{\frac{\delta}{2}-1}\frac{1}{b}$. Thus we have:

$$\mathrm{E}\Big[\frac{U_{i,1}}{U_{i,2}}\Big] \leq \mathrm{E}[U_{i,1}]\mathrm{E}\Big[\frac{1}{U_{i,2}}\Big],$$

with equality if and only if $c = ab$, that is only when $U_{i,1}$ and $U_{i,2}$ are independent.

4.2 Inference and Bayesian ratemaking

The joint distribution of $(N_{i,t}, Y_{i,t})$ given $(U_{1,i}, U_{2,i})$ has already been derived in subsection 2.5.2. From that expression and the independence Assumption 3bis, it is straightforward to deduce the likelihood function in closed form, allowing for standard maximum likelihood estimation of the model. These calculations are omitted. Let us now focus on the posterior Bayesian ratemaking. To this end we introduce the cumulative intensity $\lambda_{i,1,\bullet} = \sum_{t=1}^m \lambda_{i,1,t}$, the cumulative weighted amounts $\lambda_{i,2,\bullet} = \sum_{t=1}^m \lambda_{i,2,t} y_{i,t}$, and the total claim count $N_{i,\bullet} =$

$\sum_{t=1}^{m} N_{i,t}$. Then we have:

$$
\begin{aligned}
\mathrm{E}[Y_{i,m+1}|(N_{i,t},Y_{i,t}),t=1,\dots,m] &= \alpha \mathrm{E}\left[\frac{\lambda_{1,i,m+1}U_{1,i}}{\lambda_{2,i,m+1}U_{2,i}}\Big|(N_{i,t},Y_{i,t}),t=1,\dots,m\right] \\
&= \alpha\frac{\lambda_{1,i,m+1}}{\lambda_{2,i,m+1}}\frac{\mathrm{E}\left[\frac{U_{1,i}}{U_{2,i}}(U_1U_2^{\alpha})^{N_{i,\bullet}}\exp(-U_{1,i}\lambda_{1,i,\bullet}-U_{2,i}\lambda_{i,2,\bullet})\right]}{\mathrm{E}[(U_1U_2^{\alpha})^{N_{i,\bullet}}\exp(-U_{1,i}\lambda_{1,i,\bullet}-U_{2,i}\lambda_{i,2,\bullet})]} \\
&= \alpha\frac{\lambda_{1,i,m+1}}{\lambda_{2,i,m+1}}\frac{\partial_1^{N_{i,\bullet}+1}\partial_2^{\alpha N_{i,\bullet}-1}\mathrm{E}[\exp(-U_{1,i}\lambda_{1,i,\bullet}-U_{2,i}\lambda_{i,2,\bullet})]}{\partial_1^{N_{i,\bullet}}\partial_2^{\alpha N_{i,\bullet}}\mathrm{E}[\exp(-U_{1,i}\lambda_{1,i,\bullet}-U_{2,i}\lambda_{i,2,\bullet})]}
\end{aligned}
\tag{4.4}
$$

Using Property 2.6, we deduce that:

$$
\partial_2^{\alpha N_{i,\bullet}-1}\mathrm{E}[\exp(-U_{1,i}s-U_{2,i}t)] = \frac{\Gamma(\frac{\delta}{2}+\alpha N_{i,\bullet}-1)}{\Gamma(\frac{\delta}{2})}\frac{(b+cs)^{\alpha N_{i,\bullet}-1}}{(1+as+bs+cst)^{\frac{\delta}{2}+\alpha N_{i,\bullet}-1}} \tag{4.5}
$$

$$
\partial_2^{\alpha N_{i,\bullet}}\mathrm{E}[\exp(-U_{1,i}s-U_{2,i}t)] = \frac{\Gamma(\frac{\delta}{2}+\alpha N_{i,\bullet})}{\Gamma(\frac{\delta}{2})}\frac{(b+cs)^{\alpha N_{i,\bullet}}}{(1+as+bs+cst)^{\frac{\delta}{2}+\alpha N_{i,\bullet}}} \tag{4.6}
$$

Then it suffices to compute the $N_{i,\bullet}+1$ (resp. $N_{i,\bullet}$) th order derivative of the RHS of (4.5) (resp. (4.6)). Let us now look at some cases where this expression can be further simplified.

The independence case. If $c = ab$, that is when $U_{1,i}$ and $U_{2,i}$ are independent, the joint Laplace transform factorizes as in equation (3.7) such that the right hand side of equation (4.4) becomes:

$$
\mathrm{E}[Y_{i,m+1}|(N_{i,t},Y_{i,t}),t=1,\dots,m] = \alpha\frac{\lambda_{1,i,m+1}}{\lambda_{2,i,m+1}}\frac{a(\frac{\delta}{2}+N_{i,\bullet})}{(1+a\lambda_{1,i,\bullet})}\frac{(\frac{1}{b}+\lambda_{2,i,\bullet})}{(\frac{\delta}{2}-1+\alpha N_{i,\bullet})}. \tag{4.7}
$$

Among the two fractions on the RHS, the reader might be very familiar with the first one. Indeed, by the Poisson-Gamma conjugacy for counts, the expected value of $U_{1,i}$, *given past claim numbers only*, is: $\mathrm{E}[U_i|N_{i,t},t=1,\dots,m] = a\frac{(\frac{\delta}{2}+N_{i,\bullet})}{1+a\lambda_{1,i,\bullet}}$, and as this equality concerns the frequency part alone, it is true regardless of the dependence structure between $U_{1,i}$ and $U_{2,i}$. As for the second fraction, by the Gamma-Gamma conjugacy, it is simply the expected value of each individual claim (i.e. average claim amount), *given only the information on severity*:

$$
\mathrm{E}[C_{i,k,m+1}|\sum_{k=1}^{N_{i,t}}C_{i,k,t}=Y_{i,t},t=1,...,m] = \frac{\alpha}{\lambda_{2,i,m+1}}\frac{\frac{1}{b}+\lambda_{2,i,\bullet}}{\frac{\delta}{2}-1+\alpha N_{i,\bullet}},
$$

where the conditioning set *does not include* the information that $n_{i,\bullet}$ is also the total number of past claims and holds true whenever $U_{1,i}$ and $U_{2,i}$ are independent or not. In particular, we can check that this average claim amount *depends on* N_i, even when $U_{1,i}$ and $U_{2,i}$ are independent.

In general, when $U_{1,i}$ and $U_{2,i}$ are dependent, the RHS of equation (4.7) is clearly different from the RHS of equation (4.4). Thus from this special independence case we deduce a very important lesson on the existing two-part approach of the literature. More precisely, *the two-part model in which the posterior future claim cost is decomposed into the product between the future expected number of accidents and the future average cost is misleading as it ignores the potential dependence between $U_{1,i}$ and $U_{2,i}$. Moreover, this mis-pricing cannot be mitigated even if the past claim count $N_{i,\bullet}$ enters into the second term.*

The case without previous claims. If $N_{i,\bullet} = 0$ and hence $\lambda_{i,2,\bullet} = 0$, that is, for an individual without previous claim, the right hand side of equation (4.4) is given by equation (4.2) and we have:

$$\mathrm{E}[Y_{i,m+1}|N_{i,\bullet} = 0] = \frac{1}{\frac{\delta}{2}-1}\alpha\frac{\lambda_{1,i,m+1}}{\lambda_{2,i,m+1}}\Big[\frac{c}{(b+c\lambda_{i,1,\bullet})^2}(1+a\lambda_{i,1,\bullet}) + (\frac{\delta}{2}-1)\frac{a}{b+c\lambda_{i,1,\bullet}}\Big].$$

Clearly, if $c = ab$, then we recover the RHS of (4.7), that is the premium for the model with independent random effects. However, unfortunately, the RHS of the above equation is generically not a monotonic function of c. Nevertheless, if $\lambda_{1,i,\bullet}$ is close to zero, which is a realistic assumption if the *a priori* frequency scores $\lambda_{1,i,t}$ and the total number of observed periods m are all small, then we get an increasing function of c. That is, in this case, the (true) posterior premium is lower than when independence is assumed. On the other hand, if $\lambda_{1,i,\bullet}$ is large, then the monotonicity reverses, since we would have $\frac{c}{(b+c\lambda_{i,1,\bullet})^2} \approx \frac{1}{c\lambda^2_{1,i,\bullet}}$, and $\frac{a}{b+c\lambda_{i,1,\bullet}} \approx \frac{a}{c\lambda_{i,1,\bullet}}$, thus they are approximately decreasing in $\lambda_{i,1,\bullet}$. In this case the (true) posterior premium is higher than when independence is assumed.

The case with one previous claim. Let us now consider the case where $N_{i,\bullet} = 1$. Using Mathematica we get the ratio between the two partial derivatives in equation (4.4):

$$\frac{(1+a\lambda_{1,i,\bullet}+b\lambda_{2,i,\bullet}+c\lambda_{1,i,\bullet}\lambda_{2,i,\bullet})}{(b+c\lambda_{1,i,\bullet})^2[\alpha(ab-c)+\frac{\delta}{2}(b+c\lambda_{1,i,\bullet})(a+c\lambda_{2,i,\bullet})]}\Big[-2(\alpha-1)c(b+c\lambda_{1,i,\bullet})(a+c\lambda_{2,i,\bullet})$$
$$+\frac{(\alpha+\frac{\delta}{2})(b+c\lambda_{1,i,\bullet})^2(a+c\lambda_{2,i,\bullet})^2}{(1+a\lambda_{1,i,\bullet}+b\lambda_{2,i,\bullet}+c\lambda_{1,i,\bullet}\lambda_{2,i,\bullet})} + \frac{(\alpha-2)(\alpha-1)}{(\alpha+\frac{\delta}{2}-1)}c^2(1+a\lambda_{1,i,\bullet}+b\lambda_{2,i,\bullet}+c\lambda_{1,i,\bullet}\lambda_{2,i,\bullet})\Big].$$

This latter is a fractional function of the past weighted sum of claim cost $\lambda_{2,i,\bullet}$, with an affine denominator and a quadratic numerator. Since its coefficients are quite tedious to compute, we resort to numerical illustrations. Below we plot the trajectory of this function in $\lambda_{2,i,\bullet}$, for three set of parameters. In the following figure, we set $\lambda_{1,i,m+1} = \lambda_{2,i,m+1} = 1$, $a = 1$, $b = 1$, $c = 0.2$, $\delta = 2.1$, $\alpha = 5.5$, and $\lambda_{1,i,\bullet} = 0.1$.

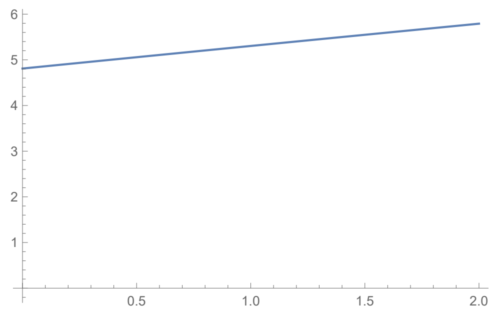

Figure 5: Plot of the posterior Bayesian expected total claim cost as a function of $\lambda_{2,i,\bullet}$.

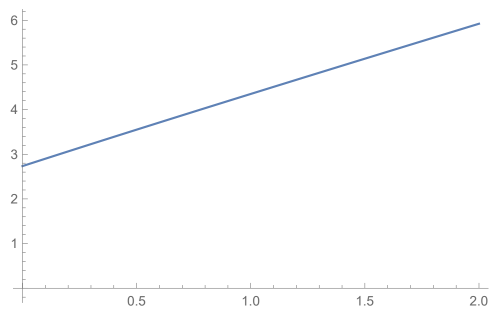

Figure 6: Plot of the same function as in the previous figure. As parameters remain the same as in Figure 5, except c is changed from 0.2 to 0.8.

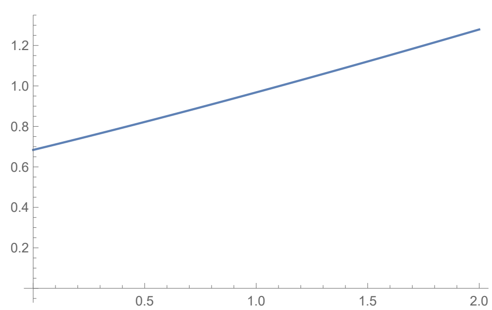

Figure 7: Plot of the same function as in the previous figure. As parameters remain the same as in Figure 5, except α is changed from 5.5 to 0.5.

In all the three plots, the posterior expectation resembles a linear function of $\lambda_{2,i,\bullet}$. Indeed, as we have said, this function is in fact fractional, and for all the three sets of parameter values, all the coefficients in the denominator and numerators are positive. For instance, for Figure 5 this function is equal to:

$$\frac{122.837 + 17.5309\lambda_{2,i,\bullet} + 0.4020\lambda_{2,i,\bullet}^2}{25.5415 + \lambda_{2,i,\bullet}} = 0.4020\lambda_{2,i,\bullet} + 7.2642 - \frac{62.70}{25.5415 + \lambda_{2,i,\bullet}},$$

which is close to linear function if the last fractional term is neglected.

The case with more than one previous claims. If $N_{i,\bullet}$ is at least equal to 2, then the expression of the posterior expectation is more tedious and will be omitted here. However, we can easily show, using equations (4.5) and (4.6), that in this case we will again have a fractional function in $\lambda_{2,i,\bullet}$, in which the denominator and numerator are polynomials of order $N_{i,\bullet}$ and $N_{i,\bullet}+1$, respectively.

4.3 Numerical illustration

Besides its tractability, another advantage of the Wishart-Gamma based bivariate random model is that under this model formulation, we can easily compare the Bayesian posterior premium with the two-part premium given by (4.7), which corresponds to the model with the same marginal distributions for the random effects but without dependence. Thus in this subsection let us compute the ratio between the two premia, and see how it differs from 1 for different parameter values.

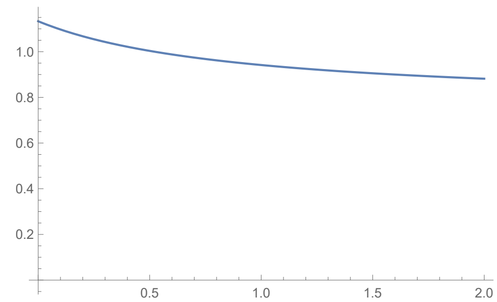

Figure 8: The Bayesian posterior expectation divided by the two-part Bayesian premium. Parameters are set to $a=1, b=1, c=0.8, \delta=2.1, \alpha=0.5, \lambda_{1,i,\bullet}=5$. Since in this case c is close to ab, the curve is close to a horizontal line with height 1.

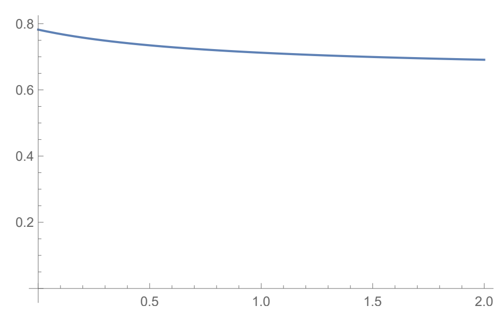

Figure 9: The Bayesian posterior expectation divided by the two-part Bayesian premium. All the parameters remain the same as in Figure 8 except that $\lambda_{1,i,\bullet}$ now takes a much smaller value 0.1 instead of 5. In other words, for new customers (with $\lambda_{1,i,\bullet}$ small), the two-part model might over-estimate the premium.

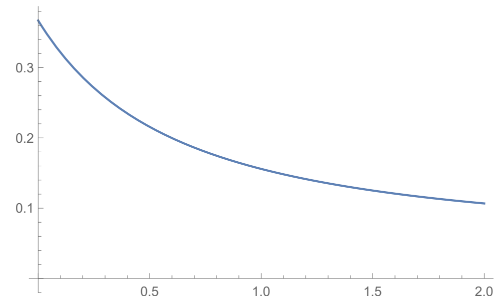

Figure 10: All the parameters remain the same as in Figure 8 except that c now takes a much smaller value 0.2 instead of 0.8, that is, in this case, the dependence between the two random effects is stronger than in the two previous models. We can see that in this case, the mispricing of the two-part model is much more important than in Figure 8.

5 Correlated time-to-event and claim amount

In the previous two sections, we have successfully shown how the Wishart-Gamma distribution can be used to build tractable mixture of (multivariate) Poisson distributions, or mixture of product distributions with Poisson and gamma margins. In this section we consider another type of applications involving duration (time-to-evet) variables. While duration variables are also positive, continuously valued just as the claim amounts, they are often observed with censoring, that is, for many individuals, the statistician might not observe directly the duration, but only the fact that it is larger than a certain threshold, which corresponds to the time elapsed up to the end of the observation period. This unique feature explains that the models and tools required to analyze such data are different from (individual or aggregate) claim amount data. Since the seminal work of Lancaster (1979), the mixed proportional hazard (MPH), which is a random effect extension of the Cox proportional hazard model, has become one of the most popular duration models [see Lu (2017) for an application to life insurance, as well as Kalbfleisch and Prentice (2011) for a textbook treatment]. In this section we will see how we can adapt the MPH methodology to a micro-reserving application involving one duration variable and one amount variable and derive closed form likelihood function and prediction formulas.

5.1 Time to payment and cash-flow in micro-reserving

To start with, let us consider claims that are settled with a single, terminal payment[2]. For claim i, denote by T_i the duration of its settlement procedure, that is, the time elapsed between the occurrence and the closure of the claim. Also, let M_i be the corresponding final amount, which only becomes observable at the settlement date. In micro-reserving, the actuary's focus is to predict the final amount for Reported but Not Settled (RBNS) claims, based on observation of settled claims. For these RBNS claims, the duration T_i is censored, that is, we only know that $T_i > C_i$, where C_i denotes the time elapsed between the

[2]In other words, even if there are potentially several payments, we only investigate the final payment as well as the cumulated final payment. This approach has recently been adopted by Lopez (2019).

occurrence of the claim and the end of the observation period, whereas the final amount M_i is to be forecasted, through the conditional expected final cost, i.e. actuaries are interested in

$$\mathrm{E}[M_i | T_i > C_i, Z_i]. \tag{5.1}$$

where Z_i is the set of observable covariates. The amount (5.1) is the case estimate for claim i. When time passes, each RBNS claim's C_i increases and the above conditional expectation needs to be updated regularly to comply with regulations.

One difficulty of dealing with censored data, especially in this bivariate setting, is that because of the dependence between T_i and M_i, the conditioning event $\{T_i > C_i\}$ cannot be omitted in (5.1). Moreover, when estimating the joint distriution of (T_i, M_i), we should take into account all the observations and not only uncensored ones, since the latter tend to have larger values for T and, assuming positive dependence between T and M, tend to have larger values for M as well. This problem has recently been tackled by Lopez (2019) using nonparametric copula estimation techniques that take into account censoring to estimate the copula function of the couple (M_i, T_i). Unfortunately this copula estimator is rather complicated, and the resulting conditional expectation involves numerical integration of the copula estimator. Let us now show that a simple, closed form estimation and forecasting procedure can be derived with the help of the Wishart-Gamma distribution.

Assumption 1ter Given $(U_{1,i}, U_{2,i}, Z_i)$, the couples (M_i, T_i) are independent for different claims i, and T_i and M_i are conditionally independent.

Assumption 2ter T_i follows a mixed proportional hazard model:

$$\mathbb{P}[T_i > t | U_{1,i}, U_{2,i}, Z_i] = \exp\Big[-\Lambda(t)\mu_{1,i}U_{1,i}\Big], \qquad \forall t > 0, \tag{5.2}$$

where $\lambda_{1,i}$ is a function of Z_i, $\Lambda(t) = \int_0^t \lambda(s)\mathrm{d}s$ is called the cumulated hazard function and the positive function $\lambda(\cdot)$ is called the baseline hazard, and the random effect $U_{1,i}$ is also called the frailty.

In particular, if $\lambda(\cdot)$ is equal to the constant function 1, then T_i is conditionally exponentially distributed with rate parameters $\mu_{1,i}U_{1,i}$. In the general case, the conditional distribution of T given U_1 is a time-changed exponential distribution, with a deterministic time change function $\Lambda(\cdot)$. In practice, the hazard function can either be specified parametrically, or non-parametrically using, say, splines. This allows T_i to have rather flexible marginal distributions.

Assumption 3ter Further, given $(U_{1,i}, U_{2,i}, Z_i)$, amount M_i is conditionally Weibull distributed:

$$\mathbb{P}[M_i > m | U_{1,i}, U_{2,i}, Z_i] = \exp\Big(- U_{2,i} \frac{m^\alpha}{\mu_{2,i}} \Big), \qquad \forall m > 0, \tag{5.3}$$

$$\text{with density: } f(m | U_{1,i}, U_{2,i}, Z_i) = \alpha m^{\alpha-1} \frac{U_{2,i}}{\mu_{2,i}} \exp\Big(- U_{2,i} \frac{m^\alpha}{\mu_{2,i}} \Big) \tag{5.4}$$

$$\text{and conditional expectation: } \mathbb{E}[m | U_{1,i}, U_{2,i}, Z_i] = \mu_{2,i}^{1/\alpha} U_{2,i}^{-1/\alpha} \Gamma(1 + \frac{1}{\alpha}). \tag{5.5}$$

Note that it would have been equally possible to specify the conditional distribution of M_i given $U_{2,i}$ as the gamma distribution with shape parameter α and rate parameter $\mu_{2,i} U_{2,i}$. Then the resulting formulas would be similar to those obtained in Section 4 and in particular are available closed form. Here we omit these similar calculation, and choose the alternative, Weibull specification. This choice is motivated by i) our aim to demonstrate the versatility of our approach; ii) the Weibull distribution has also been proposed in the bonus-malus literature to model the claim amount.

We can check that conditional survival functions (5.2) and (5.3) are decreasing functions of $U_{1,i}$ and $U_{2,i}$, respectively, thus if (U_1, U_2) are positively correlated then the correlation between T_i and M_i is positive as well. In practice, the duration and the amount are most likely to be positively correlated since claims with a long duration often involve major accident and lengthy legal process, and are therefore more costly. As a consequence, the Wishart-Gamma distribution appears to be a natural candidate in this application.

5.2 Inference

For claim i, the (multiplicative) contribution to the likelihood function of (T_i, M_i) is its joint density. This latter depends on whether the claim has been fully paid:

- if claim i has been paid after a duration T_i, then the joint density is:

$$\begin{aligned} \ell(t_i, m_i) &= \frac{\alpha m_i^{\alpha-1} \lambda(t_i) \mu_{1,i}}{\mu_{2,i}} \mathrm{E}[U_{1,i} U_{2,i} e^{-\Lambda(t_i)\mu_{1,i} U_{1,i} - \frac{m_i^\alpha}{\mu_{2,i}} U_{2,i}}] \\ &= \frac{\alpha m_i^{\alpha-1} \lambda(t_i) \mu_{1,i}}{\mu_{2,i}} \partial_1^1 \partial_2^1 \mathrm{E}[e^{-\Lambda(t_i)\mu_{1,i} U_{1,i} - \frac{m_i^\alpha}{\mu_{2,i}} U_{2,i}}], \end{aligned} \tag{5.6}$$

  which involves the $(1, \alpha)-$th order partial derivative of the joint Laplace transform of (U_1, U_2) and its closed form is given by Property 2.5.

- if instead claim i is censored by $C_i = c_i$, then the contribution to the likelihood is the probability alone:

$$\ell(t_i) = \mathbb{P}[T_i > c_i | U_{1,i}, U_{2,i}, Z_i] = \exp\Big[- \Lambda(c_i) \mu_{1,i} U_{1,i} \Big], \tag{5.7}$$

  which involves simply the marginal Laplace transform of $U_{1,i}$.

As a consequence, the (log-)likelihood function allows for closed form expression in our model.

5.3 Forecasting ultimate claim cost

Let us now compute the conditional expectation (5.1), that is the reserving amount for RBNS claims. We have, by Bayes formula:

$$\mathrm{E}[M_i|T_i > C_i] = \Gamma(1+\frac{1}{\alpha})\mu_{2,i}^{1/\alpha}\frac{\mathrm{E}\left[U_{2,i}^{-1/\alpha}e^{-\Lambda(C_i)\mu_{1,i}U_{1,i}}\right]}{\mathrm{E}[e^{-\Lambda(C_i)\mu_{1,i}U_{1,i}}]}, \tag{5.8}$$

where the numerator involves the partial fractional integral of the joint Laplace transform of (U_1, U_2), whose closed form formula is given by Property 2.6. Simple algebra leads to:

Lemma 5.1.

$$\mathrm{E}[M_i|T_i > C_i] = \mu_{2,i}^{1/\alpha}\frac{\Gamma(\frac{\delta}{2}-\frac{1}{\alpha})}{\Gamma(\frac{\delta}{2})}\Big[\frac{1+a\Lambda(C_i)\mu_{1,i}}{b+c\Lambda(C_i)\mu_{1,i}}\Big]^{1/\alpha}. \tag{5.9}$$

In particular, we need α to satisfy $\alpha > \frac{2}{\delta}$ in order for the above conditional expectation to be finite.

Let us now remind that parameters a, b, c satisfy: $c \in [0, ab]$, and in the case limiting case where $c = ab$, that is when $U_{1,i}$ and $U_{2,i}$ are independent, we have $\frac{1+a\Lambda(C_i)\mu_{1,i}}{b+c\Lambda(C_i)\mu_{1,i}} = \frac{1}{b}$ is constant. If instead $c < ab$, then we have the following Lemma:

Lemma 5.2. *The expectation $\mathrm{E}[M_i|T_i > C_i]$ is an increasing function of the elapsed time up to censoring C_i, and its domain, when C_i varies between 0 and ∞, is $[\frac{1}{b^{1/\alpha}}\mu_{2,i}^{1/\alpha}, \left(\frac{a}{c}\right)^{1/\alpha}\mu_{2,i}^{1/\alpha}]$.*

The monotonicity translates the fact that in average, the longer the settlement process, the more costly the claim. Note that for given a and b and δ, that is when the marginal distributions of $U_{1,i}$ and $U_{2,i}$ are fixed, parameter c controls the correlation between $U_{1,i}$ and $U_{2,i}$. More precisely, *ceteris parabus*, the stronger the correlation, the more variation there is for the expectation $\mathrm{E}[M_i|T_i > C_i]$ when C_i varies, or in other words the more informative the censoring variable.

Proof. It is easily checked that the function $x \mapsto \frac{1+ax}{b+cx}$ is increasing in x since its derivative is $\frac{ab-c}{(b+cx)^2} \geq 0$. Thus, since the cumulative hazard $\Lambda(C_i)$ is increasing in C_i, the RHS of equation (5.9) is increasing in C_i. When C_i goes to infinity, the ratio $\frac{1+a\Lambda(C_i)\mu_{1,i}}{b+c\Lambda(C_i)\mu_{1,i}}$ goes to $\frac{a}{c}$ when $\Lambda(C_i)$ goes to infinity (which means all claims will eventually be settled). $\square$

To illustrate the flexibility of the curve $C_i \mapsto \mathrm{E}[M_i|T_i > C_i]$, in the following figure we illustrate how it differs for different parameter values. We hold $a = b = \mu_{1,i} = \mu_{2,i} = 1$ constant, and consider three different cumulative hazard functions:

- Hazard case 1: $\Lambda(x) = x^{3/2}, \forall x > 0$. This corresponds to a convex cumulative hazard function for T_i, with increasing hazard function $\lambda(x) = \frac{\mathrm{d}}{\mathrm{d}x}\Lambda(x) = \frac{3}{2}x^{1/2}$.

- Hazard case 2: $\Lambda(x) = x, \forall x > 0$. This corresponds to a linear cumulative hazard, or constant hazard function $\lambda(x) = 1, \forall x > 0$.

- Hazard case 3: $\Lambda(x) = x^{1/2}, \forall x > 0$. This corresponds to a concav cumulative hazard function, with decreasing hazard function $\lambda(x) = \frac{1}{2}x^{-1/2}$.

We also consider two different values of α. Indeed since M_i is positive, we can also define its conditional cumulative hazard function given $U_{2,i}$, even if it is not a duration variable *per se*. From the corresponding conditional survival function given by equation (5.3), this cumulative hazard is equal to $\frac{U_{2,i}m^\alpha}{den}$, and hence α measures the convexity of the cumulative hazard function for M_i, or equivalently the monotonicity of the corresponding hazard function. Thus the two values we investigate are:

- $\alpha = 2$, corresponding to a convex cumulative conditional hazard function for $M_i|U_{2,i}$, or equivalently an increasing conditional hazard
- $\alpha = \frac{1}{2}$, corresponding to a decreasing hazard.

Finally, we choose two different values of c to illustrate the impact of dependence between $U_{1,i}$ and $U_{2,i}$ on the prediction of the claim amount.

- $c = 0.2$, which is closer to the case of comonotonicity between $U_{1,i}$ and $U_{2,i}$, that is when $c = 0$.
- $c = 0.8$, which is closer to the case of independence, that is when $c = 1$.

Below we compute the curve of $C_i \mapsto \mathrm{E}[M_i|T_i > C_i]$, for each of the three choices of the cumulative hazard function $\Lambda(x)$, and for three couples $(\alpha, c) = (0.5, 0.2)$, $(\alpha, c) = (2, 0.2)$, $(\alpha, c) = (2, 0.8)$.

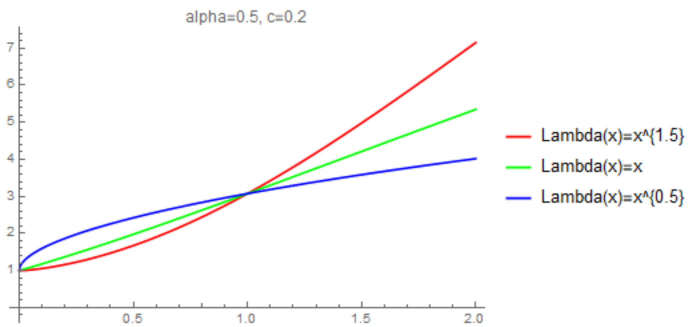

Figure 11: Curves of $C_i \mapsto \mathrm{E}[M_i|T_i > C_i]$ when $(\alpha, c) = (0.5, 0.2)$, for different choices of the cumulative hazard function Λ. In all these examples the range of $\mathrm{E}[M_i|T_i > C_i]$ is $[1, 5^2] = 1, 25[]$.

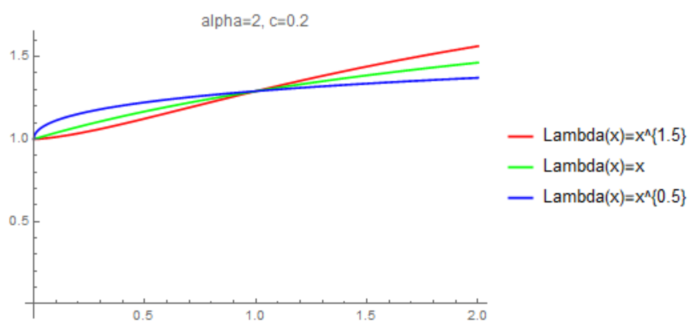

Figure 12: Curves of $C_i \mapsto \mathrm{E}[M_i|T_i > C_i]$ when $(\alpha, c) = (0.5, 0.2)$, for different choices of the cumulative hazard function Λ. In all these examples the range of $\mathrm{E}[M_i|T_i > C_i]$ is $[1, 5^{0.5}] = [1, 2.23]$

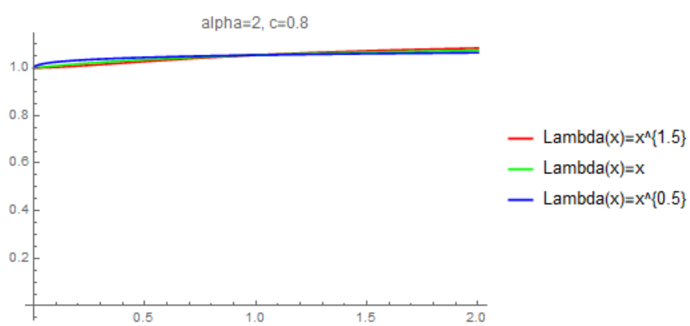

Figure 13: Curves of $C_i \mapsto \mathrm{E}[M_i|T_i > C_i]$ when $(\alpha, c) = (0.5, 0.2)$, for different choices of the cumulative hazard function Λ. In all these examples the range of $\mathrm{E}[M_i|T_i > C_i]$ is $[1, 1/0.8^{0.5}] = [1, 1.12]$.

In Figure 5, we can see that for C_i small, say between 0 and 2, the curve can be convex, nearly linear, or concave, and has the same concavity as $\Lambda(x)$. This is expected since:

$$\Big[\frac{1 + a\Lambda(C_i)\mu_{1,i}}{b + c\Lambda(C_i)\mu_{1,i}}\Big]^{1/\alpha} \approx \frac{1}{b^{1/\alpha}}\Big[1 - \frac{1}{\alpha}(a - \frac{c}{b})\Lambda(C_i)\Big], \qquad \forall C_i \approx 0.$$

Note that since the expected ultimate claim amount is upper bounded, for each choice of

$\Lambda(\cdot)$, the curve will become concave for large C_i. This is especially visible in Figure 6, where the upper bound is $5^{0.5} = 2.23$ is significantly lower than that in Figure 5. In Figure 7, however, since the range of the curve is very

6 Conclusion

In this paper we have introduced the multivariate Wishart-Gamma random effect models and demonstrated its usefulness in a large spectrum of problems encountered in general insurance, with suitable response variables ranging from count variables, amount variables to duration variables. The Wishart-Gamma distribution is a flexible extension of the univariate gamma distribution, and inherits the mathematical tractability of the latter when applied to random effect contexts.

Throughout the paper, we have made the assumption that although the random effect is multivariate, it is time invariant. Recently there have been much effort to include dynamic random effects to accommodate for possible changes in policyholder's behavior. For instance, Lu (2018) has shown that such a model is tractable, if the dynamic random effect has a marginal distribution corresponding to the multivariate Wishart-Gamma distribution introduced in this paper. The Wishart-Gamma mixture model introduced in this paper can be extended by introducing a time series with Wishart marginal distribution, called autoregressive Wishart process after Gourieroux et al. (2009). Secondly, rather than specifying the prior distribution of the dynamic random effect process, leading to a parameter-driven model, one can also investigate observation-driven, and in particular score-driven models based on Wishart distribution. Such models are gaining popularity in the time series literature (see e.g. Harvey and Fernandes, 1989, Creal et al., 2013) and actuarial science (see e.g. Bolance et al., 2007, Abdallah et al., 2016). These extensions are left for future research.

Finally, one shortcoming of the Wishart-Gamma distribution is that only positive dependence can arise between its different components. This assumption might be restrictive for certain specific applications [see e.g. Lu (2019)]. One natural idea would be to further extend the matrix Wishart distribution to its non-central version. Similar as in equation (2.1), the non-central Wishart distribution, in its simplest form, is the outer product of a Gaussian vector, but with a non-zero mean vector. It has been studied in the finance literature by Gourieroux et al. (2009) and its biggest advantage is that it still allows for closed form Laplace transform, which suggests its suitability for insurance applications as well.

Acknowledgements

Yang Lu thanks the Society of Actuaries for financial support through a CKER individual grant, the French National Center for Scientific Research CNRS through a CNRS délégation grant as well as the National Research Agency ANR through the Labex MME-DII grant. Part of the work is conducted while Lu is on sabbatical visit at THEMA, University of Cergy-Pontoise, whose hospitality is greatly acknowledged.

References

- Aalen, O.O. (1987). Mixing distributions on a Markov chain. Scandinavian Journal of Statistics 14, 281-289.

- Abdallah, A., Boucher, J.-P., Cossette, H. (2016). Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. Insurance: Mathematics and Economics 68, 120-133.

- Alai, D.H., Landsman, Z., Sherris, M. (2013). Lifetime dependence modelling using a truncated multivariate Gamma distribution. Insurance: Mathematics and Economics 52, 542-549.

- Anderson, T. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

- Antonio, K., Godecharle, E., Van Oirbeek, R. (2016). A multi-state approach and flexible payment distributions for micro-level reserving in general insurance. Available at SSRN: https://ssrn.com/abstract=2777467

- Badescu, A., Lin, X., Tang, D., Valdez, E. (2015). Multivariate Pascal mixture regression models for correlated claim frequencies. University of Toronto DP.

- Barges, M., Cossette, H., Marceau, E. (2009). TVaR-based capital allocation with copulas. Insurance: Mathematics and Economics 45, 348-361.

- Baumgartner, C., Gruber, L.F., Czado, C. (2015). Bayesian total loss estimation using shared random effects. Insurance: Mathematics and Economics 62, 194-201.

- Bermudez, L., Karlis, D. (2017). A posteriori ratemaking using bivariate Poisson models. Scandinavian Actuarial Journal 2017, 148-158.

- Bernardoff, P. (2006). Which multivariate Gamma distributions are infinitely divisible? Bernoulli 12, 169-189.

- Bolance, C., Denuit, M., Guillen, M., Lambert, Ph. (2007). Greatest accuracy credibility with dynamic heterogeneity: The Harvey-Fernandes model. Belgian Actuarial Bulletin 7, 14-18.

- Boloorforoosh, A., Christoffersen, P., Gourieroux, C., Fournier, M. (2019). Beta risk in the cross-section of equities. Forthcoming Review of Financial Studies.

- Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. Journal of Risk and Insurance 73, 687-718.

- Chiarella, C., Da Fonseca, J., Grasselli, M. (2014). Pricing range notes within Wishart affine models. Insurance: Mathematics and Economics 58, 193-203.

- Chiu, M.C., Wong, H.Y. (2014). Mean-variance asset-liability management with asset correlation risk and insurance liabilities. Insurance: Mathematics and Economics 59, 300-310.

- Cummins, J. D., Dionne, G., McDonald, J. B., and Pritchett, B. M. (1990). Applications of the GB2 family of distributions in modeling insurance loss processes. Insurance: Mathematics and Economics, 9(4):257-272.

- Creal, D., Koopman, S.J., Lucas, A. (2013). Generalized autoregressive score models with applications. Journal of Applied Econometrics 28, 777-795.

- Czado, C., Gschlossl, S. (2007). Spatial modelling of claim frequency and claim size in non-life insurance. Scandinavian Actuarial Journal 2007, 202-225.

- Czado, C., Kastenmeier, R., Brechmann, E.C., Min, A. (2012). A mixed copula model for insurance claims and claim sizes. Scandinavian Actuarial Journal, 2012(4), 278-305.

- Dionne, G., Vanasse, C. (1989). A generalization of automobile insurance rating models: The Negative Binomial distribution with a regression component. ASTIN Bulletin 19, 199-212.

- Elfving, G. (1947). A simple method of deducing certain distributions connected with multivariate sampling. Scandinavian Actuarial Journal 1947, 56-74.

- Englund, M., Guillen, M., Gustafsson, J., Nielsen, L.H., Nielsen, J.P. (2008). Multivariate latent risk: A credibility approach. ASTIN Bulletin 38, 137-146.

- Frees, E.W., Wang, P. (2005). Credibility using copulas. North American Actuarial Journal 9, 31-48.

- Furman, E., Landsman, Z. (2005). Risk capital decomposition for a multivariate dependent Gamma portfolio. Insurance: Mathematics and Economics 37, 635-649.

- Garrido, J., Genest, C., Schulz, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims. Insurance: Mathematics and Economics, 70, 205-215.

- Gourieroux, C., Jasiak, J., Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. Journal of Econometrics 150, 167-181.

- Gourieroux, C., Lu, Y. (2019). Negative Binomial autoregressive process with stochastic intensity. Journal of Time Series Analysis 40, 225-247.

- Griffiths, R. (1984). Characterization of infinitely divisible multivariate Gamma distributions. Journal of Multivariate Analysis 15, 13-20.

- Harvey, A.C., Fernandes, C. (1989). Time series models for count or qualitative observations. Journal of Business and Economic Statistics 7, 407-417.

- Henderson, R., Shimakura, S. (2003). A serially correlated Gamma frailty model for longitudinal count data. Biometrika 90, 355-366.

- Hong, L., Martin, R. (2019). Model misspecification, Bayesian versus credibility estimation, and Gibbs posteriors. Forthcoming Scandinavian Actuarial Journal.

- Jeong, H., Valdez, E.A., Ahn, J.Y., Park, S. (2017). Generalized linear mixed models for dependent compound risk models. Available at SSRN 3045360.

- Kalbfleisch, J.D., Prentice, R.L. (2011). The Statistical Analysis of Failure Time Data. Wiley, New York.

- Kramer, N., Brechmann, E.C., Silvestrini, D., Czado, C. (2013). Total loss estimation using copula-based regression models. Insurance: Mathematics and Economics 53, 829-839.

- Krishnaiah, P., Rao, M. (1961). Remarks on a multivariate Gamma distribution. The American Mathematical Monthly 68, 342-346.

- Krishnamoorthy, A., Parthasarathy, M. (1951). A multivariate Gamma-type distribution. Annals of Mathematical Statistics 22, 549-557.

- Lancaster, T. (1979). Econometric Methods for the Duration of Unemployment. Econometrica, 47(4), 939-956.

- Laue, G. (1980). Remarks on the relation between fractional moments and fractional derivatives of characteristic functions. Journal of Applied Probability 17, 456-466.

- Lee, G.Y., Shi, P. (2019). A dependent frequency–severity approach to modeling longitudinal insurance claims. Insurance: Mathematics and Economics, 87, 115-129.

- Lopez, O. (2019). A censored copula model for micro-level claim reserving. Insurance: Mathematics and Economics 87, 1-14.

- Lu, Y. (2017). Broken-Heart, Common Life, Heterogeneity: Analyzing the Spousal Mortality Dependence. ASTIN Bulletin, 47(3), 837-874.

- Lu, Y. (2018). Dynamic frailty count process in insurance: A unified framework for estimation, pricing, and forecasting. Journal of Risk and Insurance 85, 1083-1102.

- Lu, Y. (2019). Flexible (panel) regression models for bivariate count-continuous data with an insurance application. Journal of the Royal Statistical Society: Series A (Statistics in Society) 182, 1503-1521.

- Miller, K.S. (1993). An Introduction to the Fractional Calculus and Fractional Differential Equations. Wiley, New York.

- Oh, R., Shi, P., Ahn, J.Y. (2019). Bonus-Malus premiums under the dependent frequency-severity modeling. Scandinavian Actuarial Journal, in press.

- Park, S.C., Kim, J. H., Ahn, J.Y. (2018). Does hunger for bonuses drive the dependence between claim frequency and severity? Insurance: Mathematics and Economics 83, 32-46.

- Pechon, F., Trufin, J., Denuit, M. (2018). Multivariate modelling of household claim frequencies in motor third-party liability insurance. ASTIN Bulletin 48, 969-993.

- Pechon, F., Denuit, M., Trufin, J. (2019a). Multivariate modelling of multiple guarantees in motor insurance of a household. European Actuarial Journal 9, 575-602.

- Pechon, F., Denuit, M., Trufin, J. (2019b). Home and Motor insurance joined at a household level using multivariate credibility. ISBA Discussion Paper, UCLouvain.

- Pinquet, J. (1998). Designing optimal bonus-malus systems from different types of claims. ASTIN Bulletin 28, 205-220.

- Pinquet, J. (2019). Poisson models with dynamic random effects and non-negative credibilities per period. Forthcoming ASTIN Bulletin.

- Shi, P., Feng, X., Ivantsova, A. (2015). Dependent frequency–severity modeling of insurance claims. Insurance: Mathematics and Economics 64, 417-428.

- Shi, P., Valdez, E.A. (2014). Multivariate Negative Binomial models for insurance claim counts. Insurance: Mathematics and Economics 55, 18-29.

- Shi, P., Zhao, Z. (2019). Regression for copula-linked compound distributions with applications in modeling aggregate insurance claims. arXiv preprint arXiv:1910.05676.

- Sverdrup, E. (1947). Derivation of the Wishart distribution of the second order sample moments by straightforward integration of a multiple integral. Scandinavian Actuarial Journal 1947, 151-166.

- Ting Lee, M.-L. (1996). Properties and applications of the Sarmanov family of bivariate distributions. Communications in Statistics-Theory and Methods 25, 1207-1222.

- Uhlig, H. (1994). On singular Wishart and singular multivariate Beta distributions. Annals of Statistics 22, 395-405.

- Varin, C., Reid, N., Firth, D. (2011). An overview of composite likelihood methods. Statistica Sinica 21, 5-42.

- Willmot, G.E., Woo, J.-K. (2012). On the analysis of a general class of dependent risk processes. Insurance: Mathematics and Economics 51, 134-141.

- Wishart, J. (1928). The generalised product moment distribution in samples from a Normal multivariate population. Biometrika 20, 32-52.

- Yang, L., Shi, P. (2019). Multiperil rate making for property insurance using longitudinal data. Journal of the Royal Statistical Society: Series A (Statistics in Society) 182, 647-668.