Annals of Actuarial Science, Vol. 7, part 1, pp. 3–25. © Institute and Faculty of Actuaries 2012
doi:10.1017/S1748499512000140
First published online 5 September 2012

Dependence modelling in multivariate claims run-off triangles

Michael Merz
University of Hamburg, Department of Business Administration, 20146 Hamburg, Germany

Mario V. Wüthrich*
ETH Zurich, RiskLab, Department of Mathematics, 8092 Zurich, Switzerland

Enkelejd Hashorva
University of Lausanne, Department of Actuarial Science, HEC, 1015 Lausanne, Switzerland

Abstract
A central issue in claims reserving is the modelling of appropriate dependence structures. Most classical models cannot cope with this task. We define a multivariate log-normal model that allows to model both, dependence between different sub-portfolios and dependence within sub-portfolios such as claims inflation. In this model we derive closed form solutions for claims reserves and the corresponding prediction uncertainty.

Keywords
Claims reserving; Run-off triangles; Multivariate claims triangles; Accounting year effects; Calendar year effects; Claims inflation

1 Introduction and motivation

Modelling dependence structures in stochastic processes is a rather difficult task. Therefore, most stochastic models start under the restrictive assumption having independence between different risk factors, different risk modules, different portfolios, different accident years, etc. This is a common situation in general insurance claims run-off modelling where one usually assumes that claims payments of different accident years are independent (see, for instance, Mack (1993) and Hertig (1985)). Under these independence assumptions, predictors and corresponding confidence bounds are derived. Of course, one is aware of the fact that positive correlation widens these confidence bounds due to less diversification benefits. Therefore, as a next (and final) step practitioners do a "top level" correlation correction by simply multiplying the confidence bounds (obtained from the independent case) with a deterministic factor.

In the present paper we derive a comprehensive bottom-up approach for dependence modelling in general insurance claims run-offs. We choose a particular multivariate log-normal chain-ladder model and in this model we derive predictors and confidence bounds in closed form. These analytical

*Correspondence to: Mario V. Wüthrich, ETH Zurich, RiskLab, Department of Mathematics, 8092 Zurich, Switzerland. E-mail: mario.wuethrich@math.ethz.ch

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

solutions are such that they allow for *any* correlation structure. From this we can then analyze the quality of the top level correlation corrections done by practitioners. Moreover, we can calibrate correlation for risk management models which is still an open issue in Solvency 2, see European Commission (2010) and Swiss Solvency Test (2006).

There are only a few related results in the literature. Basically, there are three streams of literature on dependence modelling in general insurance claims run-offs:

(*1*) *Claims inflation and accounting year dependence within a claims run-off triangle.* The first source of dependence that has been identified by practitioners is claims inflation within a run-off triangle which acts on all accident years simultaneously. The classical distribution-free chain-ladder model of Mack (1993) cannot cope with this form of dependence. First papers that study this form of dependence are Barnett & Zehnwirth (2000), Brehm (2002), de Jong (2006), Kirschner *et al.* (2008), Kuang *et al.* (2008), Shi *et al.* (2012) and Wüthrich (2010). The first five papers mainly study this form of dependence using either maximum-likelihood methods or bootstrap simulation methods. The last two papers study accounting year dependence in a Bayesian inference framework using Markov chain Monte Carlo simulation methods. The present paper extends and combines the model and the ideas of de Jong (2006) (which goes back to Hertig (1985)) and the Bayesian inference framework studied in Shi *et al.* (2012) and Wüthrich (2010). This combination is done such that we are able to derive analytical results which allow to analyze prediction uncertainty for solvency purposes.

(*2*) *Cell-wise dependence between run-off triangles.* A second stream of literature studies the aggregation of different claims reserving triangles (sub-portfolios), see Braun (2004), Merz & Wüthrich (2007, 2008a), Shi & Frees (2011) and Zhang *et al.* (2012). These papers study a cell-wise dependence between different run-off triangles (sub-portfolios). Our model will also allow to model this form of dependence. It will turn out that this form of cell-wise dependence is a rather weak one compared to accounting year dependence.

(*3*) *Combination of cell-wise dependence between and accounting year dependence within run-off triangles.* The third stream of literature combines (*1*) and (*2*). Probably the first two papers that look at both forms of dependence in a rigorous mathematical way are Shi *et al.* (2012) and Zhang *et al.* (2012). The former develops the modelling framework using a Bayesian methodology, whereas the latter describes inflation modelling in the concluding discussion. The present paper is in the spirit of Shi *et al.* (2012). Shi *et al.* (2012) use an additive model for incremental payments where dependence is modelled in two directions (accident year and accounting year). The accounting year dependence modelling in Shi *et al.* (2012) can be interpreted as a hierarchical structure which leads to a new interesting family of models related to hierarchical generalized linear models (HGLM), see Gigante *et al.* (2012).

The starting point of our model is Hertig's log-normal chain-ladder model. We embed this multiplicative claims reserving model into a multivariate log-normal framework. This multivariate framework is chosen such that we can simultaneously model several claims run-off triangles (sub-portfolios). In this model we can choose any correlation structure between the sub-portfolios and also within the sub-portfolios. Parameter uncertainty is modelled with a multivariate Gaussian prior distribution (which is a conjugate prior to the multivariate log-normal distribution, see Bühlmann & Gisler (2005)). This combination then allows to give a closed form solution for the claims reserving problem. Moreover, it allows to derive closed form confidence bounds which, at the same time, study process uncertainty and parameter uncertainty. These results are then used to study sensitivities and, moreover, they can be used to calibrate correlation estimates for solvency purposes.

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

The nature of our multiplicative model and of the additive model of Shi *et al*. (2012) is such that they allow to utilize mathematical properties of multivariate Gaussian structures (leading to closed form solutions). The main weakness of these models is that they use a multivariate Gaussian copula as dependence structure. This dependence structure does not enjoy tail dependence (see McNeil *et al*. (2005)) and therefore joint extremes are probably under-estimated, this is also stated in Shi & Frees (2011). Moreover, often the log-normal distribution is not sufficiently heavy-tailed for modelling the tails of extremes. However, analytical formulas can only be found under these assumptions and their generalization will require simulations. Therefore, our model can serve as a benchmark for more evolved simulation-based methods. Moreover, we would like to highlight that we get fundamental findings from our model (see Conclusions 1–5 below) which explain that the state-of-the-art handling of correlation in practice should be changed, because correlation is often under-estimated (and mis-specified).

Organization of the paper. In the next section we define the multivariate Bayesian claims reserving model. In Section 3 we describe the prediction problem from a general point of view in our modelling framework. The surprisingly simple Theorem 3.1 is the key result that paves the way for closed form solutions. In Section 4 we adapt the notion from the previous section to claims reserving triangles. In Section 5 we describe the claims development result which is a crucial risk factor in Solvency 2 and we analyze prediction uncertainty in this balance sheet position. Finally, in Section 6 we present an extended case study which results in several conclusions summarized in Section 7. All proofs are provided in the appendix.

2 Model assumptions

Accident years are denoted by $i \in \{1,\dots,I\}$ and development years by $j \in \{0,\dots, J\}$. We assume that $I \geq J+1$ and that all claims are settled after development year J. Cumulative claims are denoted by $C_{i,j,n} > 0$, where the last index $n \in \{1,\dots,N\}$ corresponds to the n-th sub-portfolio (or line of business). We define the individual log-link ratios $\xi_{i,j,n}$ by (set $C_{i,-1,n} \equiv 1$ for all i and n)

$$\xi_{i,j,n} = \log\left(C_{i,j,n}/C_{i,j-1,n}\right) \quad \Leftrightarrow \quad C_{i,j,n} = C_{i,j-1,n}\exp\left\{\xi_{i,j,n}\right\}. \tag{2.1}$$

We stack the random variables $\xi_{i,j,n}$ into random vectors which will provide a handy notation. For $i \in \{1,\dots,I\}$ and $j \in \{0,\dots,J\}$ we define the vectors

$$\xi_{i,j} = (\xi_{i,j,1},\dots,\xi_{i,j,N})' \in \mathbb{R}^N, \quad \boldsymbol{\xi}_i = (\xi'_{i,0},\dots,\xi'_{i,J})' \in \mathbb{R}^a \text{ and } \boldsymbol{\xi} = (\boldsymbol{\xi}_1{}',\dots,\boldsymbol{\xi}_I{}')' \in \mathbb{R}^d,$$
with $a = N(J+1)$ and $d = aI$.

Remarks.

- The multiplicative structure (2.1) in the development period j will imply that we have a chain-ladder type model, see Mack (1993).
- The parameter $n \in \{1,\dots,N\}$ denotes the sub-portfolios (or different lines of business). If we have only one sub-portfolio, i.e. $N = 1$, then we set $\xi_{i,j,n} = \xi_{i,j}$ and we are in the classical one-dimensional claims reserving problem.
- Under the assumptions that $N = 1$ and that the $\boldsymbol{\xi}_{i,j}$'s are independent Gaussian distributed with mean θ_j we are in the framework of Hertig's log-normal claims reserving model, see Hertig (1985). Our aim is to relax these independence assumptions and to extend the model to $N > 1$.

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

Model Assumptions 2.1 (multivariate log-normal claims reserving model)

- Conditionally, given parameter $\boldsymbol{\Theta} \in \mathbb{R}^a$, the random vector $\boldsymbol{\xi}$ has a multivariate Gaussian distribution with fixed positive-definite covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ and conditional expected values $\mathbb{E}[\boldsymbol{\xi}_i|\boldsymbol{\Theta}] = \boldsymbol{\Theta}$, for all $i \in \{1,\ldots,I\}$.
- The parameter $\boldsymbol{\Theta}$ has a multivariate Gaussian distribution with prior mean $\boldsymbol{\mu} \in \mathbb{R}^a$ and positive-definite prior covariance matrix $T \in \mathbb{R}^{a\times a}$.

□

Often it is easier to describe the model assumptions by simply writing down the density function. Under Model Assumptions 2.1, the joint density of $\boldsymbol{\xi}$ and $\boldsymbol{\Theta}$ at position $(\boldsymbol{\xi},\boldsymbol{\theta})$ is given by

$$\begin{aligned} f(\boldsymbol{\xi},\boldsymbol{\theta}) = &\frac{1}{(2\pi)^{d/2}\det(\Sigma)^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{\xi}-A\boldsymbol{\theta})'\Sigma^{-1}(\boldsymbol{\xi}-A\boldsymbol{\theta})\right\} \\ &\times\frac{1}{(2\pi)^{a/2}\det(T)^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})'T^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right\}, \end{aligned} \tag{2.2}$$

with matrix $A = (1,\ldots,1)' \in \mathbb{R}^{d\times a}$ consisting of $I = d/a$ identity matrices denoted by $1 \in \mathbb{R}^{a\times a}$ and describing the conditional mean $\mathbb{E}[\boldsymbol{\xi}|\boldsymbol{\Theta}] = A\boldsymbol{\Theta} = (\boldsymbol{\Theta}',\ldots,\boldsymbol{\Theta}')' \in \mathbb{R}^d$.

Remarks 2.2

- We choose a multivariate Gaussian model for the individual log-link ratios $\xi_{i,j,n}$ which gives log-normal distributions for cumulative claims $C_{i,j,n}$. For risk management and solvency purposes one might argue that this is not a sufficiently heavy-tailed model and, moreover, this model does not have tail dependence. Nevertheless, we believe that the results from this model are interesting and important also on their own and that our model may serve as a benchmark for the ad-hoc methods used in practice and for more complex multivariate extreme value models (which still need to be developed in this field and which, in general, can only be solved numerically, for instance, using Markov chain Monte Carlo simulation methods).
- The covariance matrix Σ will allow for any correlation structure between the individual log-link ratios $\xi_{i,j,n}$ such as accounting year dependence or dependence between different sub-portfolios n. We provide explicit choices in Section 6 below.
- Parameter uncertainty is modelled through the choice of a prior distribution for $\boldsymbol{\Theta}$. Thus, a Bayesian inference analysis will include parameter estimation uncertainty in a natural way through posterior distributions.
- The model assumptions imply that conditionally, given $\boldsymbol{\Theta}$,

$$C_{1,0,n} = \exp\{\xi_{1,0,n}\},\ \ldots, C_{I,0,n} = \exp\{\xi_{I,0,n}\} \text{ are identically distributed.} \tag{2.3}$$

  That is, there is no specific accident year $i\in\{1,\ldots,I\}$ parameter under Model Assumptions 2.1. This model assumption can easily be relaxed if it is not appropriate in a particular situation. There are several different possibilities to do so, for instance

  (1) if there are known prior differences, choose suitable constants for $C_{i,-1,n} = v_{i,n} \in \mathbb{R}_+$ which give multiplicative prior adjustments, i.e. $C_{i,0,n} = v_{i,n}\exp\{\xi_{i,0,n}\}$;

  (2) a second way extends $\xi_{i,0,n} = \xi^{(1)}_{i,0,n} + \xi^{(2)}_{i,0,n}$ to a bivariate term with $\xi^{(1)}_{i,0,n}$ similar to (2.3) and independent Gaussian random variables $\xi^{(2)}_{i,0,n}$ reflecting differences between accident years i;

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

(3) or define the model conditional on the first columns, i.e. given $C_{i,0,n}$ for all i and n.

All these proposals lead to closed form solutions.
In our analysis we will calculate posterior distributions given observations in the upper claims reserving triangle $\mathcal{D}_t$, $t \geq I$, see (4.1) and (4.2) below. This will imply that the exposure of accident year i is always proportional to the last observation of accident year i at time t, i.e. proportional to $C_{i,t-i,n}$, see Theorem 4.1. This is rather close to option (3) from above and the restrictive assumption (2.3) only influences parameter estimation. Therefore, we refrain from increasing the complexity of our model.

- Crucial for the closed form solution is the multivariate Gaussian structure (2.2) which is the key property to all results in the next section. This is similar to the results in the additive model of Shi *et al.* (2012), formulas (1)–(4). Relaxation of this assumption may lead to hierarchical generalized linear models (HGLM) as they are used, for instance, in Gigante *et al.* (2012). However, many of the analytical properties get lost in these more complex models and either simulations or approximations need to be used.
- The model can be generalized to sub-portfolios n all having different dimensions I and J. The analysis will be the same but the notation will become more cumbersome, therefore we refrain from doing so.

3 Unconditional and predictive distributions

The Bayesian Model Assumptions 2.1 describe the distribution of $\boldsymbol{\xi}$, given the parameter $\boldsymbol{\Theta}$. Since, in general, this parameter is not known experts specify a prior distribution for it. In classical credibility theory one then collects observations $\xi_{i,j,n}$ and calculates the posterior distribution of $\boldsymbol{\Theta}$, given these observations, see e.g. Bühlmann & Gisler (2005). In the present situation it is simpler to directly work with the unconditional distribution of $\boldsymbol{\xi}$. That is, though we have a classical Bayesian model including parameter uncertainty, we do not describe the posterior distribution of the parameter $\boldsymbol{\Theta}$ explicitly, but we directly work on the data having the parameter distribution of $\boldsymbol{\Theta}$ as a latent factor (this can be done because the parameter $\boldsymbol{\Theta}$ acts in a simple way as a location parameter as the third proof of Theorem 3.1 will show). Thus, in a first step we calculate the unconditional distribution of $\boldsymbol{\xi}$, and in a second step we calculate the predictive distribution of $\boldsymbol{\xi}$ if we observe some of its components $\boldsymbol{\xi}_{i,j,n}$.

Theorem 3.1 (**unconditional distribution**) *Under Model Assumptions 2.1 the random vector* $\boldsymbol{\xi}$ *has a multivariate Gaussian distribution with the first two moments given by*

$$\mathbb{E}[\boldsymbol{\xi}] = A\boldsymbol{\mu} \quad \textit{and} \quad S = \mathrm{Cov}(\boldsymbol{\xi}) = \Sigma + ATA'.$$

In the proof of Theorem 3.1 we derive the following (side-)result for the covariance matrix S.

Corollary 3.2 *Under Model Assumptions 2.1 we have for the covariance matrix* $S = \Sigma + ATA'$

$$S^{-1} = \Sigma^{-1} - \Sigma^{-1}A(A'\Sigma^{-1}A + T^{-1})^{-1}A'\Sigma^{-1}.$$

Next our aim is to describe the distribution of $\boldsymbol{\xi}$ if we have observed some of its components $\boldsymbol{\xi}_{i,j,n}$. We choose a non-empty, real subset

$$\mathcal{D} \subset \{(i,j,n);\ i = 1,\dots,I,\ j = 0,\dots,J,\ n = 1,\dots,N\} \stackrel{\text{def.}}{=} \mathcal{J},$$

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

i.e. $\mathcal{D} \notin \{\mathcal{J}, \varnothing\}$. Let $|\mathcal{D}|$ denote the cardinality of $\mathcal{D}$. Define $P_{\mathcal{D}}$ to be the projection $\mathbb{R}^d \to \mathbb{R}^{|\mathcal{D}|}$ onto the components $(i,j,n) \in \mathcal{D}$, that is,

$$\boldsymbol{\xi} \mapsto \boldsymbol{\xi}^{\mathcal{D}} = P_{\mathcal{D}} \boldsymbol{\xi},$$

such that $\boldsymbol{\xi}^{\mathcal{D}}$ exactly contains the components of $\boldsymbol{\xi}$ which are in $\mathcal{D}$. Analogously, $P_{\mathcal{D}^c}$ denotes the projection onto the components in the complement $\mathcal{D}^c = \mathcal{J} \backslash \mathcal{D}$ of $\mathcal{D}$ and $\boldsymbol{\xi}^{\mathcal{D}^c}$ denotes the corresponding components. Thus, we consider the (bijective) decomposition

$$\boldsymbol{\xi} \mapsto (\boldsymbol{\xi}^{\mathcal{D}}, \boldsymbol{\xi}^{\mathcal{D}^c}) \tag{3.1}$$

which separates $\mathbb{R}^d$ into two disjoint (non-empty) subspaces $\mathbb{R}^{|\mathcal{D}|}$ and $\mathbb{R}^{|\mathcal{D}^c|}$. Our aim is to predict $\boldsymbol{\xi}^{\mathcal{D}^c}$ when we have observed $\boldsymbol{\xi}^{\mathcal{D}}$.

Lemma 3.3 *Under Model Assumptions 2.1 the random vector* $(\boldsymbol{\xi}^{\mathcal{D}}, \boldsymbol{\xi}^{\mathcal{D}^c})$ *has a multivariate Gaussian distribution with the first two moments given by*

$$\begin{aligned} \boldsymbol{\mu}_{\mathcal{D}} &= \mathbb{E}\big[\boldsymbol{\xi}^{\mathcal{D}}\big] = P_{\mathcal{D}} A \boldsymbol{\mu} \quad \text{and} \quad S_{\mathcal{D}} = \mathrm{Cov}\big(\boldsymbol{\xi}^{\mathcal{D}}\big) = P_{\mathcal{D}} S P'_{\mathcal{D}}, \\ \boldsymbol{\mu}_{\mathcal{D}^c} &= \mathbb{E}\big[\boldsymbol{\xi}^{\mathcal{D}^c}\big] = P_{\mathcal{D}^c} A \boldsymbol{\mu} \quad \text{and} \quad S_{\mathcal{D}^c} = \mathrm{Cov}\big(\boldsymbol{\xi}^{\mathcal{D}^c}\big) = P_{\mathcal{D}^c} S P'_{\mathcal{D}^c}. \end{aligned}$$

The covariance matrix between the components $\boldsymbol{\xi}^{\mathcal{D}}$ *and* $\boldsymbol{\xi}^{\mathcal{D}^c}$ *is given by*

$$S'_{\mathcal{D}^c,\mathcal{D}} = S_{\mathcal{D},\mathcal{D}^c} = \mathrm{Cov}\big(\boldsymbol{\xi}^{\mathcal{D}}, \boldsymbol{\xi}^{\mathcal{D}^c}\big) = P_{\mathcal{D}} S P'_{\mathcal{D}^c}.$$

Note that $\mathrm{Cov}(\boldsymbol{\xi})$ (with one argument) denotes the covariance matrix of the random vector $\boldsymbol{\xi}$, and, by a slight abuse of notation, we denote by $\mathrm{Cov}(\boldsymbol{\xi},\boldsymbol{\eta})$ (with two arguments) the covariance matrix between the two random vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$. We are now ready to give the crucial statement providing the predictive distribution of $\boldsymbol{\xi}^{\mathcal{D}^c}$, conditionally given observations $\boldsymbol{\xi}^{\mathcal{D}}$.

Theorem 3.4 (**predictive distribution**) *Under Model Assumptions 2.1 we have the following statement: the conditional distribution of* $\boldsymbol{\xi}^{\mathcal{D}^c}$, *given* $\boldsymbol{\xi}^{\mathcal{D}}$, *is a multivariate Gaussian distribution with conditional mean given by*

$$\boldsymbol{\mu}^{\mathrm{post}}_{\mathcal{D}^c} = \mathbb{E}\big[\boldsymbol{\xi}^{\mathcal{D}^c} \big| \boldsymbol{\xi}^{\mathcal{D}}\big] = \boldsymbol{\mu}_{\mathcal{D}^c} + S_{\mathcal{D}^c,\mathcal{D}} (S_{\mathcal{D}})^{-1} \big(\boldsymbol{\xi}^{\mathcal{D}} - \boldsymbol{\mu}_{\mathcal{D}}\big),$$

and conditional covariance matrix given by

$$S^{\mathrm{post}}_{\mathcal{D}^c} = \mathrm{Cov}\big(\boldsymbol{\xi}^{\mathcal{D}^c} \big| \boldsymbol{\xi}^{\mathcal{D}}\big) = S_{\mathcal{D}^c} - S_{\mathcal{D}^c,\mathcal{D}} (S_{\mathcal{D}})^{-1} S_{\mathcal{D},\mathcal{D}^c}.$$

4 Ultimate claim prediction and prediction uncertainty

We apply Theorem 3.4 to the claims reserving problem. Assume that we are at time $t \geq I$ (with $t < I + J$). Then, we have observed the cumulative claims $C_{i,j,n}$ with $i + j \leq t$. Note that we choose $t \geq I$ to ensure that the first column $\{(i,j,n);\ 1 \leq i \leq I,\ j = 0\}$ of each claims development triangle $n \in \{0,\ldots,N\}$ has been observed. Thus, the data is determined by the indices

$$\mathcal{D}_t = \big\{(i,j,n) \in \mathcal{J}; i + j \leq t\big\}, \tag{4.1}$$

and the resulting σ-field at time t is given by

$$\mathcal{F}_t \overset{\mathrm{def.}}{=} \sigma\big\{C_{i,j,n};\ (i,j,n) \in \mathcal{D}_t\big\} = \sigma\Big\{\xi_{i,j,n}; (i,j,n) \in \mathcal{D}_t\Big\} = \sigma\big\{\boldsymbol{\xi}^{\mathcal{D}_t}\big\}.$$

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

The Bayesian predictor for the ultimate claim $C_{i,J,n}$, $i \geq t-J+1$, at time $t \geq I$ is given by

$$\widehat{C}^{(t)}_{i,J,n} = \mathbb{E}\big[C_{i,J,n}\big|\mathcal{F}_t\big] = C_{i,t-i,n}\mathbb{E}\left[\exp\left\{\sum_{j=t-i+1}^{J} \xi_{i,j,n}\right\}\middle|\boldsymbol{\xi}^{\mathcal{D}_t}\right]. \tag{4.2}$$

Assume that $1 \leq t-i+1 \leq J$. Then we can define the set of indices

$$\mathcal{D}^c_{t|i,n} = \mathcal{D}^c_t \cap \big\{(l,j,m) \in \mathcal{J};\, l=i,\, m=n\big\},$$

which provides the projection $P_{\mathcal{D}^c_{t|i,n}} : \mathbb{R}^d \to \mathbb{R}^{J-(t-i)}$. Thus, we have

$$\mathbf{e}'_{t|i,n}\,\boldsymbol{\xi}^{\mathcal{D}^c_t} \overset{\text{def.}}{=} (1,\ldots,1)' P_{\mathcal{D}^c_{t|i,n}}\boldsymbol{\xi} = \sum_{j=t-i+1}^{J} \xi_{i,j,n}.$$

Note that this sum exactly considers all un-observed components of $\boldsymbol{\xi}^{\mathcal{D}^c_t}$ for a given accident year i and a given sub-portfolio n.

Theorem 4.1 (**ultimate claim prediction**) *Under Model Assumptions 2.1 the ultimate claim predictor* $\widehat{C}^{(t)}_{i,J,n}$ *of sub-portfolio* $n \in \{0,\ldots,N\}$ *and accident year* $i \in \{t-J+1,\ldots,I\}$ *is at accounting year* $t \in \{I,\ldots,I+J-1\}$ *given by*

$$\widehat{C}^{(t)}_{i,J,n} = C_{i,t-i,n}\exp\left\{\mathbf{e}'_{t|i,n}\,\boldsymbol{\mu}^{\text{post}}_{\mathcal{D}^c_t} + \frac{1}{2}\,\mathbf{e}'_{t|i,n}\,S^{\text{post}}_{\mathcal{D}^c_t}\,\mathbf{e}_{t|i,n}\right\}.$$

If we have incomplete data, i.e. if we have less information than $\mathcal{F}_t$ at time t, we obtain a similar result as long as the cumulative claims $C_{i,t-i,n}$ are observable at time t. For more on methods on incomplete triangles we refer to Dahms (2008).

Corollary 4.2 (**claims reserves**) *Under Model Assumptions 2.1 the claims reserves defined by* $\widehat{R}^{(t)}_{i,n} = \mathbb{E}\big[C_{i,J,n}-C_{i,t-i,n}\big|\mathcal{F}_t\big]$ *of sub-portfolio* $n \in \{0,\ldots,N\}$ *and accident year* $i \in \{t-J+1,\ldots,I\}$ *are at accounting year* $t \in \{I,\ldots,I+J-1\}$ *given by*

$$\widehat{R}^{(t)}_{i,n} = C_{i,t-i,n}\left(\exp\left\{\mathbf{e}'_{t|i,n}\,\boldsymbol{\mu}^{\text{post}}_{\mathcal{D}^c_t} + \frac{1}{2}\mathbf{e}'_{t|i,n}\,S^{\text{post}}_{\mathcal{D}^c_t}\mathbf{e}_{t|i,n}\right\}-1\right).$$

In order to analyze prediction uncertainty one typically studies the volatility of the differences

$$\widehat{C}^{(t)}_{i,J,n}-C_{i,J,n} = \widehat{R}^{(t)}_{i,n}-\big(C_{i,J,n}-C_{i,t-i,n}\big), \tag{4.3}$$

i.e. how much the true ultimate claim $C_{i,J,n}$ may deviate from its prediction $\widehat{C}^{(t)}_{i,J,n}$ at time t. Theorem 4.1 allows for a closed form calculation of the ultimate claim predictor $\widehat{C}^{(t)}_{i,J,n}$ and Theorem 3.4 allows for a simulation based analysis of any risk measure on the difference (4.3). Since we do not want to rely on simulations we choose a particular risk measure which can be calculated in closed form. The conditional mean square error of prediction (MSEP) is defined by

$$\text{msep}_{\sum_{i,n} C_{i,J,n}|\mathcal{F}_t}\left(\sum_{i,n}\widehat{C}^{(t)}_{i,J,n}\right) = \mathbb{E}\left[\left(\sum_{i,n} C_{i,J,n}-\sum_{i,n}\widehat{C}^{(t)}_{i,J,n}\right)^2\middle|\mathcal{F}_t\right].$$

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

In view of the Bayesian predictor (4.2), i.e. taking conditional expectations, we have

$$\mathrm{msep}_{\sum_{i,n} C_{i,J,n}|\mathcal{F}_t}\left(\sum_{i,n}\widehat{C}_{i,J,n}^{(t)}\right)=\mathrm{Var}\left(\sum_{i,n}C_{i,J,n}\middle|\mathcal{F}_t\right)=\sum_{i,l,n,m}\mathrm{Cov}\big(C_{i,J,n},C_{l,J,m}\big|\mathcal{F}_t\big). \tag{4.4}$$

Thus, for the conditional MSEP we need to calculate these conditional covariances.

Theorem 4.3 (**ultimate claim prediction uncertainty**) *Under Model Assumptions 2.1 we obtain for accounting year* $t\in\{I,\ldots,I+J-1\}$

$$\mathrm{msep}_{\sum_{i,n} C_{i,J,n}|\mathcal{F}_t}\left(\sum_{i,n}\widehat{C}_{i,J,n}^{(t)}\right)=\sum_{i,l,n,m}\widehat{C}_{i,J,n}^{(t)}\widehat{C}_{l,J,m}^{(t)}\left(\exp\left\{\mathbf{e}'_{t|i,n}\,S^{\mathrm{post}}_{\mathcal{D}_t^c}\mathbf{e}_{t|l,m}\right\}-1\right),$$

where the summation runs over $i,l\in\{t-J+1,\ldots,I\}$ *and* $n,m\in\{1,\ldots,N\}$.

Theorems 4.1 and 4.3 give closed form solutions for the ultimate claim prediction and the conditional MSEP analysis. The key to these results is Theorem 3.4. Moreover, Theorem 3.4 provides the full predictive distribution of the inexperienced part of the claims development which would allow to analyze any other risk measure using Monte Carlo simulations.

5 Claims development result

An important quantity in solvency considerations is the so-called claims development result (CDR). The CDR describes the changes in the predictors $\widehat{C}_{i,J,n}^{(t)}$ if we update the information from time t to $t+1$, i.e. $\mathcal{F}_t\to\mathcal{F}_{t+1}$. The CDR at time $t+1$ for accident year i and sub-portfolio n is defined by

$$CDR_{i,n}^{(t+1)}=\widehat{C}_{i,J,n}^{(t)}-\widehat{C}_{i,J,n}^{(t+1)}.$$

Because our successive Bayesian predictions (4.2) are $(\mathcal{F}_t)_t$-martingales we obtain the identity

$$\mathbb{E}\left[CDR_{i,n}^{(t+1)}\middle|\mathcal{F}_t\right]=0.$$

This explains that $CDR_{i,n}^{(t+1)}$ is typically predicted by 0 at time t, see also Merz & Wüthrich (2008b). For the conditional MSEP of this prediction we obtain the identity

$$\begin{aligned}\mathrm{msep}_{\sum_{i,n} CDR_{i,n}^{(t+1)}|\mathcal{F}_t}(0)&=\mathbb{E}\left[\left(\sum_{i,n}CDR_{i,n}^{(t+1)}-0\right)^2\middle|\mathcal{F}_t\right]=\mathrm{Var}\left(\sum_{i,n}CDR_{i,n}^{(t+1)}\middle|\mathcal{F}_t\right)\\&=\mathrm{Var}\left(\sum_{i,n}\widehat{C}_{i,J,n}^{(t+1)}\middle|\mathcal{F}_t\right)=\sum_{i,l,n,m}\mathrm{Cov}\left(\widehat{C}_{i,J,n}^{(t+1)},\widehat{C}_{l,J,m}^{(t+1)}\middle|\mathcal{F}_t\right).\end{aligned} \tag{5.1}$$

Our aim is to analyze the covariances on the right-hand side of this identity. Therefore, we need to rewrite the Bayesian ultimate claim predictors (4.2). Theorem 4.1 implies in a first step

$$\begin{aligned}\widehat{C}_{i,J,n}^{(t+1)}&=C_{i,t+1-i,n}\exp\left\{\left(\mathbf{e}'_{t+1|i,n}\boldsymbol{\mu}^{\mathrm{post}}_{\mathcal{D}_{t+1}^c}+\frac{1}{2}\mathbf{e}'_{t+1|i,n}S^{\mathrm{post}}_{\mathcal{D}_{t+1}^c}\mathbf{e}_{t+1|i,n}\right)1_{\{i+J>t+1\}}\right\}\\&=C_{i,t-i,n}\exp\left\{\boldsymbol{\xi}_{i,t+1-i,n}+\left(\mathbf{e}'_{t+1|i,n}\boldsymbol{\mu}^{\mathrm{post}}_{\mathcal{D}_{t+1}^c}+\frac{1}{2}\mathbf{e}'_{t+1|i,n}S^{\mathrm{post}}_{\mathcal{D}_{t+1}^c}\mathbf{e}_{t+1|i,n}\right)1_{\{i+J>t+1\}}\right\},\end{aligned} \tag{5.2}$$

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

where $1_{\{\cdot\}}$ denotes the indicator function. Observe that only the first two terms in the above exponent depend on the observations $\boldsymbol{\xi}^{\mathcal{D}_{t+1}}$ at time $t+1$, and this dependence is linearly as we will just show. In analogy to (3.1) we decouple (bijectively) these observations in $\mathcal{D}_{t+1}$ as follows:

$$\boldsymbol{\xi}^{\mathcal{D}_{t+1}} \mapsto (\boldsymbol{\xi}^{\mathcal{D}_t}, \boldsymbol{\xi}^{\mathcal{D}_{t+1}\setminus\mathcal{D}_t}).$$

The first term describes the components that are observable at time t and the second term the components that are observed in accounting year $t+1$, which also means $\mathcal{D}_{t+1}\setminus\mathcal{D}_t \subset \mathcal{D}_t^c$. Thus, for the latter components we also need to apply Theorem 3.4, conditionally given $\mathcal{F}_t$.

The first random term on the right-hand side of (5.2), given $\mathcal{F}_t$, is $\boldsymbol{\xi}_{i,t+1-i,n}$. We define a linear map by choosing $\mathbf{b}_{t|i,n} \in \mathbb{R}^{|\mathcal{D}_t^c|}$ such that $\mathbf{b}'_{t|i,n}\boldsymbol{\xi}^{\mathcal{D}_t^c} = \boldsymbol{\xi}_{i,t+1-i,n}$. The second random term on the right-hand side of (5.2) is given by

$$\boldsymbol{\mu}^{\text{post}}_{\mathcal{D}_{t+1}^c} = \boldsymbol{\mu}_{\mathcal{D}_{t+1}^c} + S_{\mathcal{D}_{t+1}^c,\mathcal{D}_{t+1}} \left(S_{\mathcal{D}_{t+1}}\right)^{-1} \left(\boldsymbol{\xi}^{\mathcal{D}_{t+1}} - \boldsymbol{\mu}_{\mathcal{D}_{t+1}}\right).$$

We decouple $\boldsymbol{\xi}^{\mathcal{D}_{t+1}}$. We define the linear function $B_{t+1} : \mathbb{R}^{|\mathcal{D}_{t+1}|} \to \mathbb{R}^{|\mathcal{D}_{t+1}|}$ with

$$B_{t+1}\boldsymbol{\xi}^{\mathcal{D}_{t+1}} = \left(\xi_{i,j,n} 1_{\{i+j=t+1\}}\right)'_{(i,j,n)\in\mathcal{D}_{t+1}},$$

that is, all components of $\boldsymbol{\xi}^{\mathcal{D}_{t+1}}$ are set equal to 0 if they are $\mathcal{F}_t$-measurable. This implies

$$\boldsymbol{\xi}^{\mathcal{D}_{t+1}} = (\mathbb{1} - B_{t+1})\boldsymbol{\xi}^{\mathcal{D}_{t+1}} + B_{t+1}\boldsymbol{\xi}^{\mathcal{D}_{t+1}},$$

where the first term is $\mathcal{F}_t$-measurable and the second term exactly corresponds to observations in accounting year $t+1$, i.e. to components in $\mathcal{D}_{t+1}\setminus\mathcal{D}_t \subset \mathcal{D}_t^c$.

To simplify notation we define a linear map by choosing $\mathbf{p}_{t|i,n} \in \mathbb{R}^{|\mathcal{D}_t^c|}$ such that

$$\mathbf{p}'_{t|i,n}\boldsymbol{\xi}^{\mathcal{D}_t^c} = \mathbf{b}'_{t|i,n}\boldsymbol{\xi}^{\mathcal{D}_t^c} + 1_{\{i+J>t+1\}}\mathbf{e}'_{t+1|i,n} S_{\mathcal{D}_{t+1}^c,\mathcal{D}_{t+1}} \left(S_{\mathcal{D}_{t+1}}\right)^{-1} B_{t+1}\boldsymbol{\xi}^{\mathcal{D}_{t+1}}.$$

This then allows to rewrite (5.2) as follows

$$\widehat{C}^{(t+1)}_{i,J,n} = g_{t|i,n}(\mathcal{F}_t) \exp\left\{\mathbf{p}'_{t|i,n}\boldsymbol{\xi}^{\mathcal{D}_t^c}\right\},$$

for an appropriate $\mathcal{F}_t$-measurable (constant) $g_{t|i,n}(\mathcal{F}_t)$.

Lemma 5.1 *Under Model Assumptions 2.1 we have*

$$g_{t|i,n}(\mathcal{F}_t) = C_{i,t-t,n} \exp\left\{(\mathbf{e}'_{t|i,n} - \mathbf{p}'_{t|i,n})\boldsymbol{\mu}^{\text{post}}_{\mathcal{D}_t^c} + \frac{1}{2}\mathbf{e}'_{t|i,n} S^{\text{post}}_{\mathcal{D}_t^c}\mathbf{e}_{t|i,n} - \frac{1}{2}\mathbf{p}'_{t|i,n} S^{\text{post}}_{\mathcal{D}_t^c}\mathbf{p}_{t|i,n}\right\}.$$

We are now ready to state the main theorem for the conditional MSEP of the CDR. Its single terms in (5.1) are given by the following result.

Theorem 5.2 (CDR prediction uncertainty) *Under Model Assumptions 2.1 we have for* $t\in\{I,\dots,I+J-1\}$, $i,l\in\{t-J+1,\dots,I\}$ *and* $n,m\in\{1,\dots,N\}$

$$\text{Cov}\left(\widehat{C}^{(t+1)}_{i,J,n}, \widehat{C}^{(t+1)}_{l,J,m} \middle| \mathcal{F}_t\right) = \widehat{C}^{(t)}_{i,J,n}\widehat{C}^{(t)}_{l,J,m}\left(\exp\left\{\mathbf{p}'_{t|i,n} S^{\text{post}}_{\mathcal{D}_t^c}\mathbf{p}_{t|l,m}\right\} - 1\right).$$

Before we are going to describe examples we need to slightly modify the vectors $\mathbf{p}_{t|i,n}$ in order to get handy formulas in the implementation.

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

Lemma 5.3 *Under the above assumptions we have*

$$B_{t+1}\boldsymbol{\xi}^{\mathcal{D}_{t+1}} = (P_{\mathcal{D}_{t+1}}P'_{\mathcal{D}_t^c})\,\boldsymbol{\xi}^{\mathcal{D}_t^c} \quad \text{and} \quad \mathbf{b}'_{t|i,n}\boldsymbol{\xi}^{\mathcal{D}_t^c} = \mathbf{e}'_{t|i,n}\,(P_{\mathcal{D}_t^c}P'_{\mathcal{D}_{t+1}}P_{\mathcal{D}_{t+1}}P'_{\mathcal{D}_t^c})\,\boldsymbol{\xi}^{D_t^c}.$$

This lemma has the nice consequence that we can express $\mathbf{p}'_{t|i,n}$ in terms that have already been introduced in Section 4, namely

$$\mathbf{p}'_{t|i,n} = \left(\mathbf{e}'_{t|i,n}(P_{\mathcal{D}_t^c}P'_{\mathcal{D}_{t+1}}) + 1_{\{i+J>t+1\}}\mathbf{e}'_{t+1|i,n}S_{\mathcal{D}_{t+1}^c,\mathcal{D}_{t+1}}\left(S_{\mathcal{D}_{t+1}}\right)^{-1}\right)(P_{\mathcal{D}_{t+1}}P'_{\mathcal{D}_t^c}).$$

6 Examples and sensitivities

In this section we study the influence of different choices of the covariance structure $\Sigma \in \mathbb{R}^{d\times d}$ on the resulting claims reserves $\widehat{R}_{i,n}^{(t)}$ (Theorem 4.1) and the corresponding conditional MSEP's for the ultimate claim prediction (Theorem 4.3) and for the CDR prediction (Theorem 5.2). Therefore, we revisit the multivariate data studied in Braun (2004). It consists of $N = 2$ claims reserving triangles. The first triangle ($n = 1$) describes a general liability (GL) insurance run-off (see Table 5 in the Appendix), the second triangle ($n = 2$) describes a motor third party liability (MTPL) insurance run-off (see Table 6 in the Appendix).

We make different choices for the covariance matrix Σ. First we rewrite this covariance matrix Σ in order to get a better understanding. Similar to the prior mean $\boldsymbol{\mu}$ we assume that the conditional standard deviation of $\xi_{i,j,n}$ does not depend on the accident year $i\in\{1,\ldots,I\}$, thus

$$\mathrm{Var}\left(\xi_{i,j,n}\middle|\boldsymbol{\Theta}\right)^{1/2} = \sigma_{j,n} \quad \text{for all } (i,j,n) \in \mathcal{J}.$$

We define the standard deviation vectors $\boldsymbol{\sigma}_\bullet = (\sigma_{0,1},\ldots,\sigma_{0,N},\ldots,\sigma_{J,1},\ldots,\sigma_{J,N})' \in \mathbb{R}^a$ for $\boldsymbol{\xi}_i$ and $\boldsymbol{\sigma} = A\boldsymbol{\sigma}_\bullet \in \mathbb{R}^d$ for $\boldsymbol{\xi}$. With this notation at hand we can rewrite the covariance matrix Σ as

$$\mathrm{Cov}\left(\boldsymbol{\xi}\middle|\boldsymbol{\Theta}\right) = \Sigma = \mathrm{diag}(\boldsymbol{\sigma})\,\Lambda\,\mathrm{diag}(\boldsymbol{\sigma}),$$

where $\Lambda \in \mathbb{R}^{d\times d}$ denotes the conditional correlation matrix of $\boldsymbol{\xi}$. The numerical values of the chosen standard deviation parameter σ are given in Tables 5 and 6 below. Our aim in the next sections is to examine different (positive-definite) choices of the correlation matrix Λ. For the prior uncertainty in $\boldsymbol{\Theta}$ we choose $T = \tau^2 1$ with $\tau = 1$ (this is a rather non-informative choice and the diagonal structure of T implies that prior uncertainty on the two portfolios is mutually independent).

6.1 Uncorrelated choice for Λ

As a benchmark model we consider $\Lambda = \mathbb{1}$, i.e. all components of $\boldsymbol{\xi}$ are conditionally independent, given $\boldsymbol{\Theta}$. This is a Bayesian version of Hertig's log-normal claims reserving model, see Hertig (1985). It corresponds to the model typically analyzed in practice where one assumes that payments in different accident years i are independent, conditionally given the model parameters (see also Mack (1993)). Under these model assumptions we obtain the results provided in Table 1. In this table we give the claims reserves and the corresponding conditional MSEP's, based on the information $\mathcal{F}_I$, for single accident years $i\in\{1,\ldots,14\}$ and on the line "total" for aggregated accident years $\sum_{i=1}^{14}$. We observe that the calculated total claims reserves $\sum_{i,n} R_{i,n}^{(I)}$ (aggregated over all accident years $i\in\{1,\ldots,14\}$ and all sub-portfolios $n\in\{1,2\}$) result in the value 8097,585.

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

Table 1. Uncorrelated case $\Lambda = 1$ for the aggregated GL and MTPL portfolio $\sum_{n=1,2} C_{i,J,n}$.

accident year	claims reserves	$\text{msep}^{1/2}_{\sum_n C_{i,J,n}\mid\mathcal{F}_I}\left(\sum_n \widehat{C}^{(I)}_{i,J,n}\right)$		$\text{msep}^{1/2}_{\sum_n CDR^{(I+1)}_{i,n}\mid\mathcal{F}_I}(0)$	
i	$\sum_n \widehat{R}^{(I)}_{i,n}$	absolute	in % reserves	absolute	in % reserves
1	0	0		0	
2	1,814	1,848	101.9%	1,848	101.9%
3	4,828	4,445	92.1%	4,094	84.8%
4	12,039	8,704	72.3%	6,969	57.9%
5	16,414	11,335	69.1%	7,831	47.7%
6	29,449	15,500	52.6%	10,186	34.6%
7	45,993	20,588	44.8%	11,684	25.4%
8	86,108	27,119	31.5%	13,962	16.2%
9	157,373	34,672	22.0%	20,528	13.0%
10	342,159	54,127	15.8%	33,799	9.9%
11	679,585	87,232	12.8%	58,747	8.6%
12	1280,707	152,922	11.9%	112,847	8.8%
13	2419,572	284,581	11.8%	207,872	8.6%
14	3021,544	379,975	12.6%	290,658	9.6%
$\text{Cov}^{1/2}$		217,693		151,274	
total	8097,585	556,337	6.9%	410,964	5.1%

Our first aim is to study the conditional MSEP's in this benchmark model. We have, in view of formula (4.4),

$$\begin{aligned}\text{msep}_{\sum_{i,n} C_{i,J,n}\mid\mathcal{F}_I}\left(\sum_{i,n}\widehat{C}^{(I)}_{i,J,n}\right) &= \sum_i \text{msep}_{\sum_n C_{i,J,n}\mid\mathcal{F}_I}\left(\sum_n \widehat{C}^{(I)}_{i,J,n}\right)\\ &\quad + \sum_{i\neq l}\text{Cov}\left(\sum_n C_{i,J,n}, \sum_m C_{l,J,m}\middle|\mathcal{F}_I\right).\end{aligned} \tag{6.1}$$

The first term on the right-hand side (rhs) of (6.1) is the conditional MSEP for single accident years (which is given on lines $i = 1,\ldots,14$ in Table 1). The second term on the rhs of (6.1) corresponds to the implied covariance from the simultaneous parameter uncertainty in $\boldsymbol{\Theta}$ over all accident years, that is,

$$\begin{aligned}\sum_{i\neq l}\text{Cov}\left(\sum_n C_{i,J,n}, \sum_m C_{l,J,m}\middle|\mathcal{F}_I\right) &= \sum_{i\neq l}\mathbb{E}\left[\text{Cov}\left(\sum_n C_{i,J,n}, \sum_m C_{l,J,m}\middle|\mathcal{F}_I,\boldsymbol{\Theta}\right)\middle|\mathcal{F}_I\right]\\ &\quad + \sum_{i\neq l}\text{Cov}\left(\sum_n \mathbb{E}\left[C_{i,J,n}\middle|\mathcal{F}_I,\boldsymbol{\Theta}\right], \sum_m \mathbb{E}\left[C_{l,J,m}\middle|\mathcal{F}_I,\boldsymbol{\Theta}\right]\middle|\mathcal{F}_I\right).\end{aligned}$$

The first term on the rhs is equal to zero (due to the uncorrelated choice given by $\Lambda = 1$). The square-root of the second term is given by 217,693 (and for the CDR uncertainty by 151,274, see Table 1). This allows to define the *implied average correlation between accident years* $\bar{\psi}_{\text{a.y.}}(0)$ by

$$\bar{\psi}_{\text{a.y.}}(0) \overset{\text{def.}}{=} \frac{\sum_{i\neq l}\text{Cov}\left(\sum_n C_{i,J,n}, \sum_m C_{l,J,m}\middle|\mathcal{F}_I\right)}{\sum_{i\neq l}\text{Var}\left(\sum_n C_{i,J,n}\middle|\mathcal{F}_I\right)^{1/2}\text{Var}\left(\sum_m C_{l,J,m}\middle|\mathcal{F}_I\right)^{1/2}} = 5.2\%. \tag{6.2}$$

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

Thus, the implied average correlation between accident years of the joint parameter uncertainty in $\boldsymbol{\Theta}$ is 5.2% in our example. Completely analogously we obtain for the implied average correlation between accident years for the CDR uncertainty $\bar{\psi}^{\mathrm{CDR}}_{\mathrm{a.y.}}(0) = 4.9\%$. These implied correlations play an important role in practice because typically they are specified/adjusted by experts. However, we will see that this notion can be quite controversial and even misleading.

6.2 Accounting year correlation and portfolio independence for Λ

In the spirit of Wüthrich (2010) we introduce accounting year dependence, but for the time-being we assume that the different portfolios $n \in \{1, \ldots, N\}$ are still independent. Thus, the correlation matrix Λ is chosen such that for a fixed $\rho \in [0,1)$ we have

$$\mathrm{Cov}\left(\xi_{i,j,n}, \xi_{l,k,n} \middle| \boldsymbol{\Theta}\right) = \sigma_{j,n}\sigma_{k,n}\rho 1_{\{i+j=l+k\}} \quad \text{for } (i,j) \neq (l,k), \tag{6.3}$$

$$\mathrm{Cov}\left(\xi_{i,j,n}, \xi_{l,k,m} \middle| \boldsymbol{\Theta}\right) = 0 \quad \text{for } n \neq m. \tag{6.4}$$

The results for these choices (6.3)–(6.4) are given in Table 2. The first observation is that the claims reserves are a decreasing function in ρ. Though very important, this observation should not be over-stated because depending on the numerical example we may also obtain an increasing function in ρ for the claims reserves. This decrease of the claims reserves is related to the fact that the data does not satisfy assumption (2.3) and we should adjust for this using prior differences as described in Remarks 2.2. To fully understand this adjustment more analysis on the micro-level of the data needs to be done. We refrain from doing so here because we would like to concentrate on the uncertainties given by the conditional MSEP's.

A very important observation is that the conditional MSEP's are an increasing function in ρ. For example, a correlation parameter of $\rho = 20\%$ leads to a substantial increase of the square-rooted CDR uncertainty of

$$\frac{533,055}{410,964} - 1 \approx 30\%!$$

Table 2. Accounting year correlation ρ and portfolio independence according to (6.3)–(6.4).

	claims reserves	$\mathrm{msep}^{1/2}_{\sum_{i,n} C_{i,J,n}\mid\mathcal{F}_I}\left(\sum_{i,n}\widehat{C}^{(I)}_{i,J,n}\right)$		$\mathrm{msep}^{1/2}_{\sum_{i,n} CDR^{(I+1)}_{i,n}\mid\mathcal{F}_I}(0)$	
accounting year correlation ρ	$\sum_{i,n}\widehat{R}^{(I)}_{i,n}$	absolute	in % reserves	absolute	in % reserves
0%	8097,585	556,337	6.9%	410,964	5.1%
10%	7842,470	650,963	8.3%	477,437	6.1%
20%	7688,058	728,723	9.5%	533,055	6.9%
30%	7572,026	796,104	10.5%	581,649	7.7%
40%	7477,039	856,212	11.5%	625,190	8.4%
50%	7395,570	910,834	12.3%	664,860	9.0%
60%	7323,574	961,113	13.1%	701,435	9.6%
70%	7258,587	1007,835	13.9%	735,455	10.1%
80%	7198,991	1051,565	14.6%	767,314	10.7%
90%	7143,657	1092,728	15.3%	797,313	11.2%

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

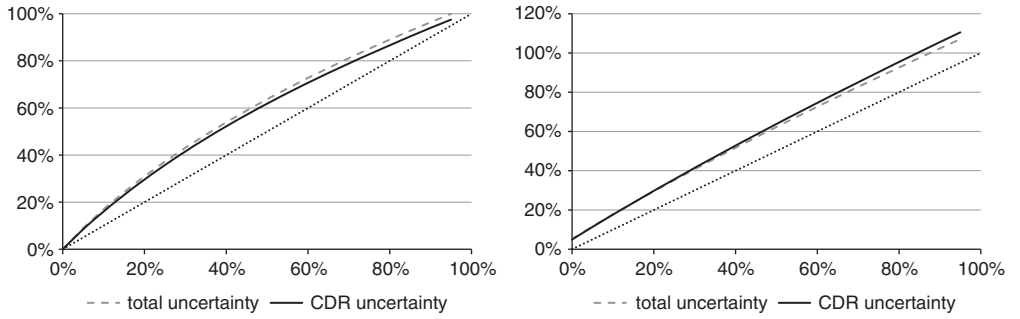


Figure 1. Accounting year correlation ρ and portfolio independence according to (6.3)–(6.4): (a) lhs: relative increase of square-rooted uncertainties as a function of the accounting year correlation $\rho \in [0,1)$ for $\mathrm{msep}^{1/2}_{\sum_{i,n} C_{i,J,n}|\mathcal{F}_I}\left(\sum_{i,n} \widehat{C}^{(I)}_{i,J,n}\right)$ and $\mathrm{msep}^{1/2}_{\sum_{i,n} CDR^{(I+1)}_{i,n}|\mathcal{F}_I}(0)$; (b) rhs: average implied correlations between accident years $\bar{\psi}_{\text{a.y.}}(\rho)$ and $\bar{\psi}^{\text{CDR}}_{\text{a.y.}}(\rho)$ as a function of the accounting year correlation $\rho \in [0,1)$.

The full curve is given in Figure 1(a) (lhs) where we see that this increase is a non-linear function that needs to be analyzed in a careful bottom-up approach.

Conclusion 1 *The introduction of accounting year dependence may substantially increase the prediction uncertainty. This relative increase is non-linear in the correlation parameter* $\rho \in [0,1)$ (Figure 1(a), *lhs*).

Next we calculate the average implied correlations between accident years $\bar{\psi}_{\text{a.y.}}(\rho)$ and $\bar{\psi}^{\text{CDR}}_{\text{a.y.}}(\rho)$ as a function of the accounting year correlation $\rho \in [0,1)$ (see also (6.2)). We define these average implied correlations between accident years by

$$\bar{\psi}_{\text{a.y.}}(\rho) \overset{\text{def.}}{=} \frac{\sum_{i \neq l} \mathrm{Cov}^{(\rho)}\left(\sum_n C_{i,J,n}, \sum_m C_{l,J,m} \middle| \mathcal{F}_I\right)}{\sum_{i \neq l} \mathrm{Var}^{(0)}\left(\sum_n C_{i,J,n} \middle| \mathcal{F}_I\right)^{1/2} \mathrm{Var}^{(0)}\left(\sum_m C_{l,J,m} \middle| \mathcal{F}_I\right)^{1/2}}, \tag{6.5}$$

where $\mathrm{Cov}^{(\rho)}(\cdot,\cdot|\mathcal{F}_I)$ corresponds to the choice $\rho \in [0,1)$ and $\mathrm{Var}^{(0)}(\cdot|\mathcal{F}_I)$ to the choice $\rho = 0$. The results are presented in Figure 1(b) (rhs). We see that $\bar{\psi}_{\text{a.y.}}(\rho)$ and $\bar{\psi}^{\text{CDR}}_{\text{a.y.}}(\rho)$ are non-linearly increasing functions in ρ. More surprisingly, we see that $\bar{\psi}_{\text{a.y.}}(\rho) > 1$ for $\rho > 85\%$ which seems to be an error at first sight. However, these figures are correct and there results a counter-intuitive correlation bigger than 1! The reason for this value bigger than 1 is that (6.5) does not give a well-specified correlation measure because enumerator and denominator belong to different models (specifications of ρ). For different ρ's also the conditional MSEP's for single accident years i (given in Table 1 for $\rho = 0$) will change; and thus Table 1 does not allow to reconstruct the results for $\rho > 0$ using an average implied correlation factor (bounded by 1). We conclude again that only a bottom-up calculation can provide the full flavour of the results, because if we need to reconstruct Table 1 for $\rho > 0$ then we directly obtain the full picture of the overall uncertainties.

Conclusion 2 *The introduction of an average implied correlation between accident years given by* (6.5) *does not allow to construct the conditional MSEP's for* $\rho > 0$ *from the conditional MSEP's with* $\rho = 0$ *in an intuitive way, because* $\bar{\psi}_{\text{a.y.}}(\rho)$ *can take arbitrary and counter-intuitive values* (*Figure 1(b), rhs*).

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

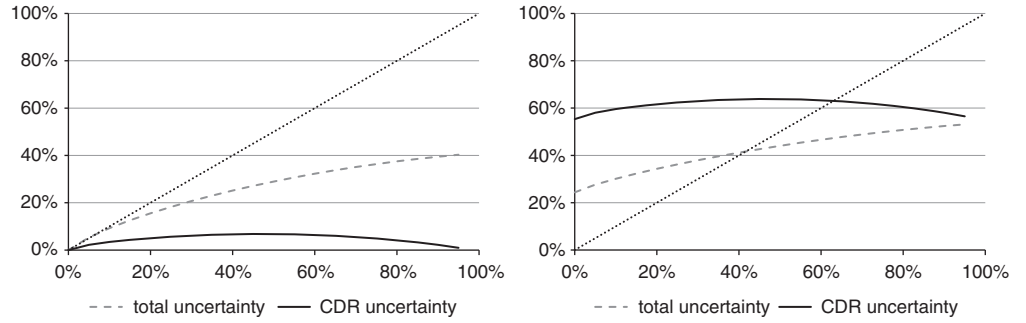


Figure 2. Example Saluz *et al.* (2011), Table 7: Accounting year correlation ρ according to (6.3): (a) lhs: relative increase of square-rooted uncertainties as a function of the accounting year correlation $\rho \in [0,1)$ for $\mathrm{msep}^{1/2}_{\sum_i C_{i,J,1}|\mathcal{F}_I}\left(\sum_i \widehat{C}^{(I)}_{i,J,1}\right)$ and $\mathrm{msep}^{1/2}_{\sum_i CDR^{(I+1)}_{i,1}|\mathcal{F}_I}(0)$; (b) rhs: average implied correlations between accident years $\bar{\psi}_{\text{a.y.}}(\rho)$ and $\bar{\psi}^{CDR}_{\text{a.y.}}(\rho)$ as a function of the accounting year correlation $\rho \in [0,1)$.

The next observation on the average implied correlation between accident years came as a surprise to us (and at first glance it seemed very counter-intuitive). To see this next phenomenon we need to choose a different data set. We choose the MTPL data from Saluz *et al.* (2011), Table 7 (and we set $N = 1$). Figure 2 shows that the square-rooted conditional MSEP $\mathrm{msep}^{1/2}_{\sum_i C_{i,J,1}|\mathcal{F}_I}\left(\sum_i \widehat{C}^{(I)}_{i,J,1}\right)$ of the total ultimate claim prediction and the corresponding implied average correlation $\bar{\psi}_{\text{a.y.}}(\rho)$ are increasing functions in the accounting year correlation ρ. The picture for the CDR is very different: in this example the conditional MSEP $\mathrm{msep}_{\sum_i CDR^{(I+1)}_{i,1}|\mathcal{F}_I}(0)$ is *not* an increasing function of the accounting year correlation parameter ρ! There are two arguments for this non-monotonicity property: (1) the ultimate claim predictor $\widehat{C}^{(I)}_{i,J,1}$ is in this example a decreasing function in ρ (similar to Table 2) and this decrease *cannot* be compensated by an increase of the covariance term $(\exp\{\mathbf{p}'_{I|i,1}\, S^{\text{post}}_{\mathcal{D}^c_I}\, \mathbf{p}_{I|l,1}\}-1)$, see Theorem 5.2; (2) the CDR uncertainty measures possible changes over the next accounting year whereas the ultimate claim prediction uncertainty measures the total uncertainty until all claims are settled. An increasing accounting year correlation ρ may also imply that more parameter uncertainty is shifted from the next year to latter periods, and hence the one-year CDR uncertainty (for the next accounting year) is relatively decreasing compared to the total uncertainty.

Conclusion 3 *The average implied correlation between accident years is not necessarily an increasing function in the accounting year correlation parameter ρ (Figure 2(b), rhs).*

6.3 Accounting year independence and portfolio correlation for Λ

In this section we study portfolio correlation in the spirit of Braun (2004) and Merz & Wüthrich (2007, 2008a), but we assume that there is no accounting year correlation. Thus, the correlation matrix Λ is chosen such that for a fixed $\rho \in [0,1)$ we have

$$\mathrm{Cov}\left(\xi_{i,j,n}, \xi_{l,k,m} \middle| \boldsymbol{\Theta}\right) = \sigma_{j,n}\sigma_{k,m}\rho 1_{\{(i,j)=(l,k)\}} \quad \text{for } (i,j,n) \neq (l,k,m), \tag{6.6}$$

i.e. this is a point-wise cell dependence between $\boldsymbol{\xi}_{i,j,n}$ and $\boldsymbol{\xi}_{i,j,m}$. The results for these choices (6.6) are given in Table 3. We observe that the increase in prediction uncertainty is comparably small under

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

Table 3. Point-wise cell correlation ρ and accounting year independence according to (6.6).

	claims reserves	$\text{msep}^{1/2}_{\sum_{i,n} C_{i,J,n}\mid\mathcal{F}_I}\left(\sum_{i,n}\widehat{C}^{(I)}_{i,J,n}\right)$		$\text{msep}^{1/2}_{\sum_{i,n} CDR^{(I+1)}_{i,n}\mid\mathcal{F}_I}(0)$	
point-wise cell correlation ρ	$\sum_{i,n}\widehat{R}^{(I)}_{i,n}$	absolute	in % reserves	absolute	in % reserves
0%	8097,585	556,337	6.9%	410,964	5.1%
10%	8097,585	575,262	7.1%	425,923	5.3%
20%	8097,585	593,605	7.3%	440,384	5.4%
30%	8097,584	611,417	7.6%	454,396	5.6%
40%	8097,584	628,743	7.8%	467,998	5.8%
50%	8097,584	645,624	8.0%	481,225	5.9%
60%	8097,584	662,092	8.2%	494,108	6.1%
70%	8097,583	678,178	8.4%	506,673	6.3%
80%	8097,583	693,909	8.6%	518,942	6.4%
90%	8097,583	709,308	8.8%	530,937	6.6%

assumption (6.6) (compare Tables 2 and 3). Under assumption (6.6) the relative increase in the square-rooted conditional MSEP can be at most $\sqrt{2}-1=41.4\%$, whereas under assumptions (6.3)–(6.6) we can get relative increases of 100% (see Figure 1(a), lhs). Thus, if we only apply this type of point-wise cell correlation, then prediction uncertainty is typically under-estimated because data show more dependence structure.

Conclusion 4 *Point-wise cell dependence only allows for a moderate increase in the uncertainty estimates.*

Conclusions 1 and 4 are in line with the findings in Shi *et al.* (2012), Table 2.

6.4 Accounting year and portfolio correlation for Λ

Finally, we combine the two previous forms of dependence. We choose Σ such that for $\rho\in[0,1)$

$$\text{Cov}\left(\xi_{i,j,n},\xi_{l,k,m}\middle|\boldsymbol{\Theta}\right) = \sigma_{j,n}\sigma_{k,m}\,\rho 1_{\{i+j=l+k\}} \quad \text{for } (i,j,n)\neq(l,k,m). \tag{6.7}$$

The results for these choices (6.7) are given in Table 4. We see a substantial increase in the uncertainty if we combine these two different types of dependence. E.g., if we choose $\rho=20\%$ we obtain the following relative increases in the CDR uncertainty:

model (6.3)–(6.4):	533,055/410,964–1 $\approx$ 30%,	see Table 2,
model (6.6):	440,384/410,964–1 $\approx$ 7%,	see Table 3,
model (6.7):	620,545/410,964–1 $\approx$ 51%,	see Table 4.

That is, already a "comparably moderate" correlation parameter of $\rho=20\%$ implies in model (6.7) an increase of CDR uncertainty of 51% (measured in the square-rooted conditional MSEP). If we start from the uncorrelated model $\rho=0$ we obtain portfolio uncertainties

$$\text{msep}^{1/2}_{\sum_i CDR^{(I+1)}_{i,n}\mid\mathcal{F}_I}(0) = \begin{cases} 372,690 & \text{for } n=1, \\ 172,613 & \text{for } n=2. \end{cases}$$

A "correlation" factor of 1 between these two latter figures implies a total uncertainty of

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

Table 4. Accounting year and portfolio correlation according to (6.7).

	claims reserves	$\text{msep}^{1/2}_{\sum_{i,n} C_{i,J,n}\mid\mathcal{F}_I}\left(\sum_{i,n}\widehat{C}^{(I)}_{i,J,n}\right)$		$\text{msep}^{1/2}_{\sum_{i,n} CDR^{(I+1)}_{i,n}\mid\mathcal{F}_I}(0)$	
correlation ρ	$\sum_{i,n}\widehat{R}^{(I)}_{i,n}$	absolute	in % reserves	absolute	in % reserves
0%	8097,585	556,337	6.9%	410,964	5.1%
10%	7707,647	718,021	9.3%	528,494	6.9%
20%	7533,921	843,029	11.2%	620,545	8.2%
30%	7421,690	948,802	12.8%	698,663	9.4%
40%	7340,040	1042,186	14.2%	767,687	10.5%
50%	7276,834	1126,728	15.5%	830,183	11.4%
60%	7225,968	1204,555	16.7%	887,706	12.3%
70%	7183,917	1277,057	17.8%	941,278	13.1%
80%	7148,458	1345,205	18.8%	991,616	13.9%
90%	7118,096	1409,706	19.8%	1039,243	14.6%

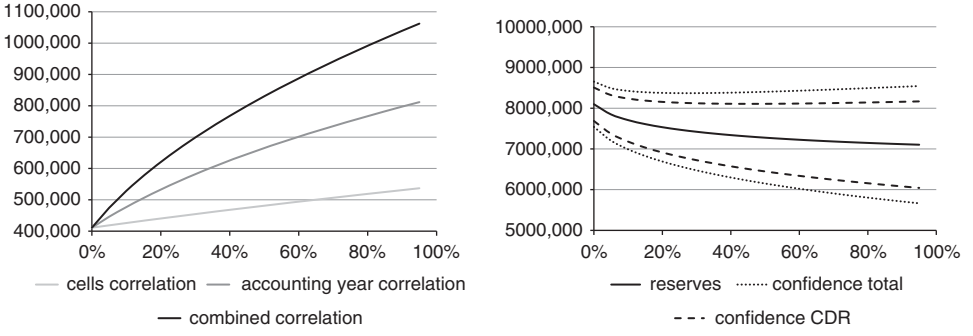


Figure 3. (a) lhs: comparison of models (6.3)–(6.4) (accounting year correlation), (6.6) (point-wise cells correlation) and (6.7) (combined correlation), we display the square-rooted conditional MSEPs of the CDR uncertainty as a function of $\rho \in [0,1)$; (b) rhs: combined correlation model (6.7), claims reserves and 1-standard deviation confidence interval for the total uncertainty and the CDR uncertainty as a function of $\rho \in [0,1)$.

372,690 + 172,613 = 545,569 which is by far less than the 620,545 from model (6.7) with $\rho = 20\%$. Thus, we get the following conclusion.

Conclusion 5 *An overall top correlation as specified in Solvency 2 (see European Commission (2010), SCR.9.34) or in the Swiss Solvency Test (see Swiss Solvency Test (2006), Section 8.4) typically under-estimates the prediction uncertainty if one starts with the benchmark model $\rho = 0$.*

Finally, in Figure 3(a) (lhs) we compare the square-rooted conditional MSEPs for the CDR uncertainties as a function of ρ for the three models considered above. We observe the non-linear increase which can only be calculated in a bottom-up approach (as suggested above). Figure 3(b)

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

(rhs) gives the claims reserves from Corollary 4.2 together with a confidence interval of 1-standard deviation, both for the total uncertainty (Theorem 4.3) and the CDR uncertainty (Theorem 5.2) as a function of the correlation parameter ρ.

7 Conclusions

We summarize the findings and give a brief outlook:

- There is no reasonable way (short-cut) to choose correlation on a top level because often intuition fails as Conclusions 1–5 show. Overall correlations can *only* be determined in a bottom-up approach.
- We provide the whole tool kit (Theorems 4.3 and 5.2) for this bottom-up approach in the multivariate log-normal model. Other distributional models can often only be solved numerically and then the sensitivity analysis becomes much intransparent.
- Other risk measures than the conditional MSEP (such as Value-at-Risk or Tail-Value-at-Risk) can only be calculated numerically. In addition, it would be interesting to analyze the behaviour of these risk measures under dependence structures with tail dependence, see McNeil *et al.* (2005).
- The choice of the (positive-definite) covariance matrix Σ needs a thorough discussion. Sometimes data is helpful but we believe that expert opinion is as important as data in order to get reasonable choices for Σ. Often accounting year dependence is correlated with inflation that depends on financial market developments. Similar to König *et al.* (2011) and Donnelly & Wüthrich (2012) these dependence structures between insurance claims and financial market movements should be studied in order to enhance the quality of the prediction.
- Sensitivities for different lines of business with different development patterns (and uncertainty) should be investigated in a comprehensive case study.

References

Barnett, G. & Zehnwirth, B. (2000). Best estimates for reserves. *Proceedings CAS*, **LXXXVII**, 245–321.

Boyd, S. & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Braun, C. (2004). The prediction error of the chain ladder method applied to correlated runoff triangles. *ASTIN Bulletin*, **34/2**, 399–423.

Brehm, P.J. (2002). Correlation and the aggregation of unpaid loss distributions. *CAS Forum*, 1–23.

Bühlmann, H. & Gisler, A. (2005). *A Course in Credibility Theory and its Applications*. Springer.

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

de Jong, P. (2006). Forecasting runoff triangles. *North American Actuarial Journal*, **10/2**, 28–38.

Donnelly, C. & Wüthrich, M.V. (2012). Bayesian prediction of disability insurance frequencies using economic factors. *Annals of Actuarial Science*, **6/2**, 381–400.

European Commission (2010). QIS5 Technical Specifications, Annex to Call for Advice from CEIOPS on QIS5.

Gigante, P., Picech, L. & Sigalotti, L. (2012). Claims reserving in the hierarchical generalized linear models framework. Preprint.

Hertig, J. (1985). A statistical approach to the IBNR-reserves in marine insurance. *ASTIN Bulletin*, **15/2**, 171–183.

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

Johnson, R.A. & Wichern, D.W. (1988). *Applied Multivariate Statistical Analysis*, 2nd edition. Prentice-Hall.
Kirschner, G.S., Kerley, C. & Isaacs, B. (2008). Two approaches to calculating correlated reserves indications across multiple lines of business. *Variance*, **2/1**, 15–38.
König, B., Weber, F. & Wüthrich, M.V. (2011). Prediction of disability frequencies in life insurance. *Zavarovalniški horizonti, Journal of Slovensko aktuarsko združenje*, **7/3**, 5–23.
Kuang, D., Nielsen, B. & Nielsen, J.P. (2008). Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika*, **95**, 987–991.
Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, **23/2**, 213–225.
McNeil, A.J., Frey, R. & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.
Merz, M. & Wüthrich, M.V. (2007). Prediction error of the chain ladder reserving method applied to correlated run-off triangles. *Annals of Actuarial Science*, **2/1**, 25–50.
Merz, M. & Wüthrich, M.V. (2008a). Prediction error of the multivariate chain ladder reserving method. *North American Actuarial Journal*, **12/2**, 175–197.
Merz, M. & Wüthrich, M.V. (2008b). Modelling the claims development result for solvency purposes. *CAS E-Forum*, **Fall 2008**, 542–568.
Pestman, W.R. (1998). *Mathematical Statistics*. de Gruyter.
Saluz, A., Gisler, A. & Wüthrich, M.V. (2011). Development pattern and prediction error for the stochastic Bornhuetter-Ferguson claims reserving method. *ASTIN Bulletin*, **41/2**, 279–313.
Shi, P., Basu, S. & Meyers, G.G. (2012). A Bayesian log-normal model for multivariate loss reserving. *North American Actuarial Journal*, **16/1**, 29–51.
Shi, P. & Frees, E.W. (2011). Dependent loss reserving using copulas. *ASTIN Bulletin*, **41/2**, 449–486.
Swiss Solvency Test (2006). FINMA SST Technisches Dokument, Version 2. October 2006.
Wüthrich, M.V. (2010). Accounting year effects modelling in the stochastic chain ladder reserving method. *North American Actuarial Journal*, **14/2**, 235–255.
Zhang, Y., Dukic, V. & Guszcza, J. (2012). A Bayesian non-linear model for forecasting insurance loss payments. *Journal Royal Statistical Society A*, **175/2**, 637–656.

Appendix A Proofs of the statements

Proof of Theorem 3.1. We give three different proofs for Theorem 3.1 because all of them are interesting on their own.

Proof (1). The first proof goes by "brute force". It uses classical Bayesian thinking that tries to determine the form of the posterior density explicitly. This proof is important because it shows how the model can be generalized to non-Gaussian model assumptions and how Markov chain Monte Carlo simulation methods need to be applied. We rewrite the joint density (2.2) of $\boldsymbol{\xi}$ and $\boldsymbol{\Theta}$ as follows (hiding all normalizing constants in the symbol $\propto$)

$$\begin{aligned}-2\log f(\boldsymbol{\xi},\boldsymbol{\theta}) &\propto (\boldsymbol{\xi}-A\boldsymbol{\theta})'\Sigma^{-1}(\boldsymbol{\xi}-A\boldsymbol{\theta}) + (\boldsymbol{\theta}-\boldsymbol{\mu})'T^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}) \\ &\propto \boldsymbol{\xi}'\Sigma^{-1}\boldsymbol{\xi}-2\boldsymbol{\theta}'A'\Sigma^{-1}\boldsymbol{\xi} + \boldsymbol{\theta}'A'\Sigma^{-1}A\boldsymbol{\theta} + \boldsymbol{\theta}'T^{-1}\boldsymbol{\theta}-2\boldsymbol{\theta}'T^{-1}\boldsymbol{\mu} \\ &= \boldsymbol{\xi}'\Sigma^{-1}\boldsymbol{\xi}-2\boldsymbol{\theta}'\left(A'\Sigma^{-1}\boldsymbol{\xi} + T^{-1}\boldsymbol{\mu}\right) + \boldsymbol{\theta}'\left(A'\Sigma^{-1}A + T^{-1}\right)\boldsymbol{\theta}.\end{aligned}$$

We define the matrices $B^{-1} = A'\Sigma^{-1}A + T^{-1} \in \mathbb{R}^{a\times a}$ and $C = A'\Sigma^{-1}A \in \mathbb{R}^{a\times a}$. These matrices are positive-definite: choose $\boldsymbol{\theta} \neq (0,\ldots,0)' \in \mathbb{R}^a$. The choice of A implies $A\boldsymbol{\theta} \neq (0,\ldots,0)' \in \mathbb{R}^d$. Since Σ is positive-definite we obtain for all these vectors $\boldsymbol{\theta}$

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

$$\boldsymbol{\theta}' C\boldsymbol{\theta} = \boldsymbol{\theta}'\left(A'\Sigma^{-1}A\right)\boldsymbol{\theta} = (A\boldsymbol{\theta})'\Sigma^{-1}A\boldsymbol{\theta} > 0,$$

thus the matrix C is positive-definite. Moreover, T is positive-definite and hence $B^{-1} = C + T^{-1}$ is positive-definite (and invertible).

We add and subtract the same term so that we obtain a density as a function of $\boldsymbol{\theta}$.

$$\begin{aligned}
-2\log f(\boldsymbol{\xi},\boldsymbol{\theta}) &\propto \boldsymbol{\xi}'\Sigma^{-1}\boldsymbol{\xi} - 2\boldsymbol{\theta}'B^{-1}B\left(A'\Sigma^{-1}\boldsymbol{\xi} + T^{-1}\boldsymbol{\mu}\right) + \boldsymbol{\theta}'B^{-1}\boldsymbol{\theta} \\
&= \boldsymbol{\xi}'\Sigma^{-1}\boldsymbol{\xi} - \left(A'\Sigma^{-1}\boldsymbol{\xi} + T^{-1}\boldsymbol{\mu}\right)'B'B^{-1}B\left(A'\Sigma^{-1}\boldsymbol{\xi} + T^{-1}\boldsymbol{\mu}\right) \\
&\quad + \left(\boldsymbol{\theta} - B\left(A'\Sigma^{-1}\boldsymbol{\xi} + T^{-1}\boldsymbol{\mu}\right)\right)'B^{-1}\left(\boldsymbol{\theta} - B\left(A'\Sigma^{-1}\boldsymbol{\xi} + T^{-1}\boldsymbol{\mu}\right)\right) \\
&\propto \boldsymbol{\xi}'\Sigma^{-1}\boldsymbol{\xi} - \boldsymbol{\xi}'\Sigma^{-1}AB'A'\Sigma^{-1}\boldsymbol{\xi} - 2\boldsymbol{\xi}'\Sigma^{-1}AB'T^{-1}\boldsymbol{\mu} \\
&\quad + \left(\boldsymbol{\theta} - B\left(A'\Sigma^{-1}\boldsymbol{\xi} + T^{-1}\boldsymbol{\mu}\right)\right)'B^{-1}\left(\boldsymbol{\theta} - B\left(A'\Sigma^{-1}\boldsymbol{\xi} + T^{-1}\boldsymbol{\mu}\right)\right).
\end{aligned}$$

The last line provides a multivariate Gaussian density for $\boldsymbol{\theta}$. This implies, note that $B' = B$,

$$f(\boldsymbol{\xi}) = \int_{\mathbb{R}^a} f(\boldsymbol{\xi},\boldsymbol{\theta})d\boldsymbol{\theta} \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\xi}'\Sigma^{-1}\boldsymbol{\xi} - \boldsymbol{\xi}'\Sigma^{-1}ABA'\Sigma^{-1}\boldsymbol{\xi} - 2\boldsymbol{\xi}'\Sigma^{-1}ABT^{-1}\boldsymbol{\mu}\right)\right\}. \tag{A.1}$$

Thus, we obtain a multivariate Gaussian distribution for $\boldsymbol{\xi}$ and there remains to calculate its mean and variance. The unconditional mean of $\boldsymbol{\xi}$ is given by

$$\mathbb{E}[\boldsymbol{\xi}] = \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\xi}|\boldsymbol{\Theta}\right]\right] = \mathbb{E}[A\boldsymbol{\Theta}] = A\boldsymbol{\mu}.$$

For the unconditional covariance matrix of $\boldsymbol{\xi}$ we have

$$\mathrm{Cov}(\boldsymbol{\xi}) = \mathbb{E}\left[\mathrm{Cov}\left(\boldsymbol{\xi}|\boldsymbol{\Theta}\right)\right] + \mathrm{Cov}\left(\mathbb{E}\left[\boldsymbol{\xi}|\boldsymbol{\Theta}\right]\right) = \Sigma + \mathrm{Cov}(A\boldsymbol{\Theta}) = \Sigma + ATA'.$$

This finishes the first proof of Theorem 3.1.

Proof (2). An elegant proof goes via the moment generating function. Choose $\mathbf{y} \in \mathbb{R}^d$. We have

$$\begin{aligned}
\mathbb{E}\left[\exp\left\{\mathbf{y}'\boldsymbol{\xi}\right\}\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp\left\{\mathbf{y}'\boldsymbol{\xi}\right\}|\boldsymbol{\Theta}\right]\right] = \mathbb{E}\left[\exp\left\{\mathbf{y}'A\boldsymbol{\Theta} + \mathbf{y}'\Sigma\mathbf{y}/2\right\}\right] \\
&= \exp\left\{\mathbf{y}'A\boldsymbol{\mu} + \mathbf{y}'ATA'\mathbf{y}/2 + \mathbf{y}'\Sigma\mathbf{y}/2\right\} = \exp\left\{\mathbf{y}'A\boldsymbol{\mu} + \mathbf{y}'(\Sigma + ATA')\mathbf{y}/2\right\}.
\end{aligned}$$

This is the moment generating function of a multivariate Gaussian distribution with the required first two moments. This finishes the second proof of Theorem 3.1.

Proof (3). This proof uses that the parameter $\boldsymbol{\Theta}$ acts as a location parameter. Assume that $\mathbf{Y}$ is independent of $\boldsymbol{\Theta}$ having a centered multivariate Gaussian distribution with covariance matrix Σ. Then we have the distributional identity

$$\boldsymbol{\xi} = \boldsymbol{\xi} - A\boldsymbol{\Theta} + A\boldsymbol{\Theta} \overset{(d)}{=} \mathbf{Y} + A\boldsymbol{\Theta}.$$

The rhs now describes a multivariate Gaussian distribution with the required first two moments. Note that the matrix ATA' is positive-semi-definite and hence S is positive-definite. This finishes the third proof of Theorem 3.1.

□

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

Proof of Corollary 3.2. From (A.1) we immediately see that the inverse covariance matrix S is given by

$$\begin{aligned} S^{-1} &= \Sigma^{-1} - \Sigma^{-1}ABA'\Sigma^{-1} = \Sigma^{-1} - \Sigma^{-1}A(C + T^{-1})^{-1}A'\Sigma^{-1} \\ &= \Sigma^{-1} - \Sigma^{-1}A(A'\Sigma^{-1}A + T^{-1})^{-1}A'\Sigma^{-1}. \end{aligned}$$

For completeness we also transform the last term in (A.1). Applying A from the right to the equality above we obtain, using the positive-definiteness of C,

$$\begin{aligned} S^{-1}A &= \Sigma^{-1}A\Big(\mathbb{1} - (C + T^{-1})^{-1}C\Big) = \Sigma^{-1}A\Big(\mathbb{1} - (C + T^{-1})^{-1}(C^{-1})^{-1}\Big) \\ &= \Sigma^{-1}A\Big(\mathbb{1} - (\mathbb{1} + C^{-1}T^{-1})^{-1}\Big) = \Sigma^{-1}A\Big(\mathbb{1} - ((T + C^{-1})T^{-1})^{-1}\Big) \\ &= \Sigma^{-1}A\Big(\mathbb{1} - T(T + C^{-1})^{-1}\Big). \end{aligned}$$

Let us calculate the last bracket, it provides

$$\begin{aligned} \mathbb{1} - T(T + C^{-1})^{-1} &= (T + C^{-1})(T + C^{-1})^{-1} - T(T + C^{-1})^{-1} \\ &= ((T + C^{-1}) - T)(T + C^{-1})^{-1} = C^{-1}(T + C^{-1})^{-1} \\ &= (TC + \mathbb{1})^{-1} = (C + T^{-1})^{-1}T^{-1} = BT^{-1}. \end{aligned}$$

We have just proved that $S^{-1}A = \Sigma^{-1}ABT^{-1}$. Thus, (A.1) becomes

$$f(\boldsymbol{\xi}) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\xi}'S^{-1}\boldsymbol{\xi} - 2\boldsymbol{\xi}'S^{-1}A\boldsymbol{\mu}\right)\right\} \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\xi} - A\boldsymbol{\mu})S^{-1}(\boldsymbol{\xi} - A\boldsymbol{\mu})\right\}.$$

This completes the proof.

□

Proof of Lemma 3.3. The lemma is an immediate consequence of Theorem 3.1. Note that the linear, bijective map (3.1) only describes a relabeling of the components (permutation) and this does not change the type of the underlying distribution. The first two moments are then also a consequence of Theorem 3.1 using the linearity of the projections, see also Result 4.4 in Johnson & Wichern (1988).

□

Proof of Theorem 3.4. In view of Lemma 3.3 this is a well-known statement for multivariate Gaussian distributions, see Pestman (1998), pages 427–428, or Result 4.6 in Johnson & Wichern (1988). The matrix $\mathrm{Cov}(\boldsymbol{\xi}^{\mathcal{D}^c}|\boldsymbol{\xi}^{\mathcal{D}})$ describes the Schur complement of $S_{\mathcal{D}}$ in S, we also refer to Section A.5.5 in Boyd & Vandenberghe (2004).

□

Proofs of Theorems 4.1 and 4.3. Note that $\mathbf{e}'_{t|i,n}\boldsymbol{\xi}^{\mathcal{D}_t^c}$ is a sum over components in $\boldsymbol{\xi}^{\mathcal{D}_t^c}$. From Theorem 3.4 we obtain that $\mathbf{e}'_{t|i,n}\boldsymbol{\xi}^{\mathcal{D}_t^c}|\boldsymbol{\xi}^{\mathcal{D}_t}$ has a Gaussian distribution with posterior mean $\mathbf{e}'_{t|i,n}\boldsymbol{\mu}^{\mathrm{post}}_{\mathcal{D}_t^c}$ and posterior variance $\mathbf{e}'_{t|i,n}S^{\mathrm{post}}_{\mathcal{D}_t^c}\mathbf{e}_{t|i,n}$. Then the claims immediately follow from the properties of log-normal distributions.

□

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

Proof of Lemma 5.1. Because successive predictions are $(\mathcal{F}_t)_t$-martingales and in view of Theorem 3.4 we have

$$\widehat{C}^{(t)}_{i,J,n} = \mathbb{E}\Big[\widehat{C}^{(t+1)}_{i,J,n}\Big|\mathcal{F}_t\Big] = g_{t|i,n}(\mathcal{F}_t)\exp\Big\{\mathbf{p}'_{t|i,n}\boldsymbol{\mu}^{\text{post}}_{\mathcal{D}^c_t} + \frac{1}{2}\mathbf{p}'_{t|i,n}S^{\text{post}}_{\mathcal{D}^c_t}\mathbf{p}_{t|i,n}\Big\}.$$

The rhs now needs to be equal to the term given in Theorem 4.1 which proves the claim.

□

Proof of Theorem 5.2. A straightforward calculation shows

$$\begin{aligned}\text{Cov}\Big(\widehat{C}^{(t+1)}_{i,J,n}, \widehat{C}^{(t+1)}_{l,J,m}\Big|\mathcal{F}_t\Big) &= \mathbb{E}\Big[\widehat{C}^{(t+1)}_{i,J,n}\widehat{C}^{(t+1)}_{l,J,m}\Big|\mathcal{F}_t\Big] - \mathbb{E}\Big[\widehat{C}^{(t+1)}_{i,J,n}\Big|\mathcal{F}_t\Big]\mathbb{E}\Big[\widehat{C}^{(t+1)}_{l,J,m}\Big|\mathcal{F}_t\Big]\\ &= g_{t|i,n}(\mathcal{F}_t)g_{t|l,m}(\mathcal{F}_t)\big(\mathbb{E}\big[\exp\big\{(\mathbf{p}'_{t|i,n} + \mathbf{p}'_{t|l,m})\boldsymbol{\xi}^{\mathcal{D}^c_t}\big\}\big|\mathcal{F}_t\big]\\ &\quad - \mathbb{E}\big[\exp\big\{\mathbf{p}'_{t|i,n}\boldsymbol{\xi}^{\mathcal{D}^c_t}\big\}\big|\mathcal{F}_t\big]\mathbb{E}\big[\exp\big\{\mathbf{p}'_{t|l,m}\boldsymbol{\xi}^{\mathcal{D}^c_t}\big\}\big|\mathcal{F}_t\big]\big)\\ &= \widehat{C}^{(t)}_{i,J,n}\widehat{C}^{(t)}_{l,J,m}\Big(\exp\Big\{\mathbf{p}'_{t|i,n}S^{\text{post}}_{\mathcal{D}^c_t}\mathbf{p}_{t|l,m}\Big\} - 1\Big).\end{aligned}$$

This proves the theorem.

□

Proof of Lemma 5.3. For the first claim we note that $\boldsymbol{\xi}^{\mathcal{D}^c_t} = P_{\mathcal{D}^c_t}\boldsymbol{\xi}$. Thus, the linear map $P_{\mathcal{D}^c_t} : \mathbb{R}^d \to \mathbb{R}^{|\mathcal{D}^c_t|}$ selects all components of $\boldsymbol{\xi}$ which are in $\mathcal{D}^c_t$. Next, we observe that the linear map $P'_{\mathcal{D}^c_t} : \mathbb{R}^{|\mathcal{D}^c_t|} \to \mathbb{R}^d$ maps the vector $\boldsymbol{\xi}^{\mathcal{D}^c_t}$ to the vector

$$\tilde{\boldsymbol{\xi}} = \big(\xi_{i,j,n}1_{\{(i,j,n)\in\mathcal{D}^c_t\}}\big)'_{(i,j,n)\in\mathcal{J}} \in \mathbb{R}^d,$$

i.e. all components $(i,j,n) \notin \mathcal{D}^c_t$ are set equal to 0. This implies that the linear map $(P'_{\mathcal{D}^c_t}P_{\mathcal{D}^c_t}) : \mathbb{R}^d \to \mathbb{R}^d$ exactly selects all the components that are in $\mathcal{D}^c_t$, i.e.

$$(P'_{\mathcal{D}^c_t}P_{\mathcal{D}^c_t})\boldsymbol{\xi} = \tilde{\boldsymbol{\xi}} = \big(\xi_{i,j,n}1_{\{(i,j,n)\in\mathcal{D}^c_t\}}\big)'_{(i,j,n)\in\mathcal{J}} \in \mathbb{R}^d.$$

If we finally apply the linear map $P_{\mathcal{D}_{t+1}} : \mathbb{R}^d \to \mathbb{R}^{|\mathcal{D}_{t+1}|}$ we see that

$$(P_{\mathcal{D}_{t+1}}P'_{\mathcal{D}^c_t}P_{\mathcal{D}^c_t})\boldsymbol{\xi} = P_{\mathcal{D}_{t+1}}\tilde{\boldsymbol{\xi}} = \Big(\xi_{i,j,n}1_{\{(i,j,n)\in\mathcal{D}^c_t\}}\Big)'_{(i,j,n)\in\mathcal{D}_{t+1}} \in \mathbb{R}^{|\mathcal{D}_{t+1}|},$$

thus, we keep the values of the components $(i,j,n) \in \mathcal{D}_{t+1} \cap \mathcal{D}^c_t$ and all components in $\mathcal{D}_{t+1}\backslash\mathcal{D}^c_t$ are set equal to 0 which exactly is equal to $B_{t+1}\boldsymbol{\xi}^{\mathcal{D}_{t+1}}$.

For the second claim we observe (using the first statement) that

$$(P'_{\mathcal{D}_{t+1}}P_{\mathcal{D}_{t+1}}P'_{\mathcal{D}^c_t})\boldsymbol{\xi}^{\mathcal{D}^c_t} = P'_{\mathcal{D}_{t+1}}B_{t+1}\xi^{\mathcal{D}_{t+1}} = \Big(\xi_{i,j,n}1_{\{(i,j,n)\in\mathcal{D}^c_t\cap\mathcal{D}_{t+1}\}}\Big)'_{(i,j,n)\in\mathcal{J}} \in \mathbb{R}^d.$$

This implies

$$(P_{\mathcal{D}^c_t}P'_{\mathcal{D}_{t+1}}P_{\mathcal{D}_{t+1}}P'_{\mathcal{D}^c_t})\boldsymbol{\xi}^{\mathcal{D}^c_t} = P_{\mathcal{D}^c_t}\Big(\xi_{i,j,n}1_{\{(i,j,n)\in\mathcal{D}_{t+1}\}}\Big)'_{(i,j,n)\in\mathcal{D}^c_t} \in \mathbb{R}^{|\mathcal{D}^c_t|}.$$

Thus, $(P_{\mathcal{D}^c_t}P'_{\mathcal{D}_{t+1}}P_{\mathcal{D}_{t+1}}P'_{\mathcal{D}^c_t})\boldsymbol{\xi}^{\mathcal{D}^c_t}$ considers all components of $\boldsymbol{\xi}^{\mathcal{D}^c_t}$ which become observable at time

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

Table 5. Observed general liability (GL) data $C_{i,j,n}$ with $i+j \leq I = 14$ and $n = 1$. The vector $\boldsymbol{\mu}_n = (\mu_{0,n}, \ldots, \mu_{J,n})'$ denotes the prior mean of parameter vector $\boldsymbol{\Theta}$ that corresponds to sub-portfolio $n = 1$, $\boldsymbol{\sigma}_n = (\sigma_{0,n}, \ldots, \sigma_{J,n})'$ denotes the standard deviation parameters implied from $\boldsymbol{\Sigma}$ for sub-portfolio $n = 1$, and for the prior covariance matrix we choose $T = \tau^2 \mathbb{1}$ with $\tau = 1$.

i/j	0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	59,966	163,152	254,512	349,524	433,265	475,778	513,660	520,309	527,978	539,039	537,301	540,873	547,696	549,589
2	49,685	153,344	272,936	383,349	458,791	503,358	532,615	551,437	555,792	556,671	560,844	563,571	562,795	
3	51,914	170,048	319,204	425,029	503,999	544,769	559,475	577,425	588,342	590,985	601,296	602,710		
4	84,937	273,183	407,318	547,288	621,738	687,139	736,304	757,440	758,036	782,084	784,632			
5	98,921	278,329	448,530	561,691	641,332	721,696	742,110	752,434	768,638	768,373				
6	71,708	245,587	416,882	560,958	654,652	726,813	768,358	793,603	811,100					
7	92,350	285,507	466,214	620,030	741,226	827,979	873,526	896,728						
8	95,731	313,144	553,702	755,978	857,859	962,825	1022,241							
9	97,518	343,218	575,441	769,017	934,103	1019,303								
10	173,686	459,416	722,336	955,335	1141,750									
11	139,821	436,958	809,926	1174,196										
12	154,965	528,080	1032,684											
13	196,124	772,971												
14	204,325													
$\boldsymbol{\mu}_n$	11.5266	1.1562	0.5312	0.2998	0.1657	0.1000	0.0532	0.0257	0.0142	0.0115	0.0062	0.0046	0.0056	0.0035
$\boldsymbol{\sigma}_n$	0.4721	0.1094	0.0700	0.0500	0.0298	0.0200	0.0150	0.0100	0.0100	0.0100	0.0080	0.0070	0.0050	0.0021

https://doi.org/10.1017/S1748499512000140 Published online by Cambridge University Press

Table 6. Observed motor third party liability (MTPL) data $C_{i,j,n}$ with $i+j \leq I = 14$ and $n = 2$. The vector $\boldsymbol{\mu}_n = (\mu_{0,n}, \ldots, \mu_{J,n})'$ denotes the prior mean of parameter vector Θ that corresponds to sub-portfolio $n = 2$, $\boldsymbol{\sigma}_n = (\sigma_{0,n}, \ldots, \sigma_{J,n})'$ denotes the standard deviation parameters implied from Σ for sub-portfolio $n = 2$, and for the prior covariance matrix we choose $T = \tau^2 1$ with $\tau = 1$.

i/j	0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	114,423	247,961	312,982	344,340	371,479	371,102	380,991	385,468	385,152	392,260	391,225	391,328	391,537	391,428
2	152,296	305,175	376,613	418,299	440,308	465,623	473,584	478,427	478,314	479,907	480,755	485,138	483,974	
3	144,325	307,244	413,609	464,041	519,265	527,216	535,450	536,859	538,920	539,589	539,765	540,742		
4	145,904	307,636	387,094	433,736	463,120	478,931	482,529	488,056	485,572	486,034	485,016			
5	170,333	341,501	434,102	470,329	482,201	500,961	504,141	507,679	508,627	507,752				
6	189,643	361,123	446,857	508,083	526,562	540,118	547,641	549,605	549,693					
7	179,022	396,224	497,304	553,487	581,849	611,640	622,884	635,452						
8	205,908	416,047	520,444	565,721	600,609	630,802	648,365							
9	210,951	426,429	525,047	587,893	640,328	663,152								
10	213,426	509,222	649,433	731,692	790,901									
11	249,508	580,010	722,136	844,159										
12	258,425	686,012	915,109											
13	368,762	909,066												
14	394,997													
$\boldsymbol{\mu}_n$	12.2147	0.7772	0.2361	0.1106	0.0639	0.0335	0.0165	0.0095	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\boldsymbol{\sigma}_n$	0.3499	0.0966	0.0500	0.0250	0.0200	0.0150	0.0100	0.0080	0.0060	0.0050	0.0040	0.0030	0.0021	0.0010