KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Business and Economics

Individual Loss Reserving with the Multivariate Skew Normal Distribution

Mathieu Pigeon, Katrien Antonio, Michel Denuit

DEPARTMENT OF ACCOUNTANCY, FINANCE AND INSURANCE (AFI)

AFI_1261

Individual Loss Reserving with the Multivariate Skew Normal Distribution

Mathieu Pigeon[*] Katrien Antonio[†] Michel Denuit[‡]

December 14, 2012

Abstract

The evaluation of future cash flows and solvency capital recently gained importance in general insurance. To assist in this process, our paper proposes a novel loss reserving model, designed for individual claims in discrete time. We model the occurrence of claims, as well as their reporting delay, the time to the first payment, and the cash flows in the development process. Our approach uses development factors similar to those of the well–known chain–ladder method. We suggest the Multivariate Skew Normal distribution as a suitable framework for modeling the multivariate distribution of development factors. Empirical analysis using a realistic portfolio and out–of–sample prediction tests demonstrate the relevance of the model proposed.

Keywords: Stochastic loss reserving, general insurance, Multivariate Skew Normal distribution, chain–ladder, individual claims.

1 Introduction

We develop a novel stochastic model for loss reserving in general insurance. The model uses detailed information on the development of individual claims. A vector of discrete random variables describes the claim's evolution over time, which evolves from occurrence of the accident till settlement or censoring of the claim. The corresponding stream of payments is expressed in terms of chain–ladder alike development factors (or: link ratios) and modeled with a multivariate, parametric distribution. The model

[*] Université Catholique de Louvain (UCL, Belgium), Mathieu.Pigeon@uclouvain.be
[†] KU Leuven (Belgium) and University of Amsterdam (UvA, The Netherlands), K.Antonio@uva.nl
[‡] Université Catholique de Louvain (UCL, Belgium), Michel.Denuit@uclouvain.be

leads to a theoretical expression for the best estimate of the outstanding amount for each claim, and a corresponding predictive distribution follows by simulation.

We divide the time structure of a general insurance claim in three parts (see Figure 1). Between occurrence of the accident and notification to the insurance company, the insurer is liable for the claim amount but is unaware of the claim's existence. The claim is said to be Incurred But Not Reported (IBNR). After notification, the claim is known by the company and the first payment (if any) will follow. In this paper, we use the expression Reported But Not Paid (RBNP) to describe an incurred and reported claim for which no payments have been made yet. Then, the initial payment occurs and several partial payments (and refunds) follow. The claim finally closes at the closure or settlement date. From the first payment to the closure of the claim, the insurer is aware of the existence of the claim, but the final amount is still unknown: the claim is Reported But Not Settled (RBNS). This structure provides a flexible framework which can be simplified or extended if necessary.

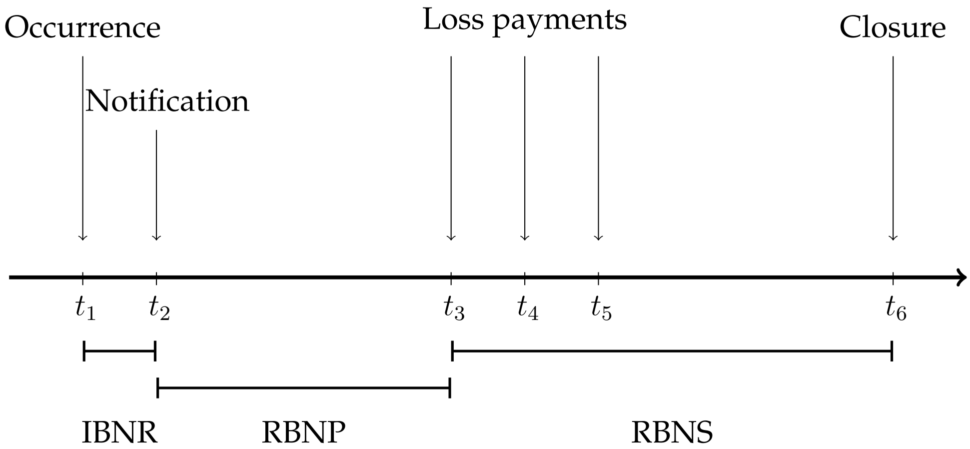


Figure 1: Evolution of a general insurance claim.

At the evaluation date the actuary should estimate technical provisions. Loosely speaking, the insurer must predict, with maximum accuracy, the total amount needed to pay claims that he has legally committed to cover. One part of the total amount comes from the completion of Reported But Not Settled (RBNS) claims. Predictions for costs related to Reported But Not Paid (RBNP) claims and Incurred But Not Reported (IBNR) claims form the second part of the total amount.

With the introduction of Solvency 2 and IFRS 4 Phase 2, the evaluation of future cash flows and regulatory required solvency capital becomes more important and current techniques for loss reserving may have to be improved, adjusted or extended. In general, existing methods for claims reserving are designed for aggregated data, conveniently summarized in a run–off triangle with occurrence and development years.

The chain–ladder approach (Mack's model in Mack (1993) and Mack (1999)) is the most popular member of this category. A rich literature exists about those techniques, see England and Verrall (2002) or Wüthrich and Merz (2008) for an overview.

Leaving the track of data aggregated in run–off triangles, Arjas (1989), Norberg (1993) and Norberg (1999) develop a mathematical framework for the development of individual claims in continuous time. More recent contributions in this direction are Zhao et al. (2009) and Antonio and Plat (2012). Verrall et al. (2010), Martinez et al. (2011) and Martinez et al. (2012) extend the traditional chain–ladder framework towards the use of extra data sources. Their work connects the triangular approach with the idea of micro–level loss reserving. We develop a model at the confluence of Norberg (1993), Antonio and Plat (2012) and the chain–ladder model. Instead of using a continuous time line, we use discrete random variables – at the level of an individual claim – for the reporting delay, the first payment delay, the number of payments and the number of periods between two consecutive payments. Individual development factors structure the development pattern, which is similar to the chain–ladder method. We propose the framework of Multivariate Skew Symmetric distributions (more specifically: the Multivariate Skew Normal distribution) to model the resulting dependent development factors at individual claim level.

Our paper is organized as follows. We introduce the statistical model in Section 2. We present the data in Section 3 and this real example is developped in Sections 4 and 5. Finally, we conclude in Section 6. Some technical developments are gathered in an appendix, for the sake of completeness.

2 The Model

Suppose we have a data set at our disposal with detailed information about the development of individual claims. More specifically, the model uses the occurrence date, the declaration date, the date(s) of payment(s) (and refund(s)) done for the claim, the amount(s) paid for the claim and the closure date.

2.1 Model Specification

2.1.1 Time Components

We denote the k^{th} claim from occurrence period i (with $k = 1, \ldots, K_i$ and $i = 1, \ldots, I$) with (ik). In our discrete framework we identify:

- the random variable T_{ik} is the *reporting delay* for claim (ik), i.e. the difference between the occurrence period of the claim and the period of its notification to the insurance company;

- the random variable Q_{ik} is the *first payment delay*, representing the difference between the notification period and the first period with payment for claim (ik);
- the random variable U_{ik} models the *number of period(s)* with partial payment after the first one; and
- the random variable N_{ikj} represents the *delay between two periods with payment* which is the number of periods between payments j and $j+1$. We use $N_{ik,U_{ik}+1}$ to denote the number of periods between the last payment and the settlement of the claim. Consequently, $N_{ik} = \sum_{j=1}^{U_{ik}+1} N_{ikj}$ is the number of periods between occurrence period and settlement of the claim.

Each component follows a discrete distribution $f : \mathbb{N} \to [0,1]$, respectively $f_1(t;\boldsymbol{\nu})$, $f_2(q;\boldsymbol{\psi})$, $f_3(u;\boldsymbol{\beta})$ and $f_4(n;\boldsymbol{\phi})$. By definition, $\Pr(N_{ikj} = 0) = 0$, $\forall j$. An example of this structure for a real–life data set is in Section 3. In the sequel of the text we will interpret 'periods' as years.

2.1.2 Exposure and Occurrence Measures

To distinguish explicitly between IBNR and RBNS/RBNP claims, we need a stochastic process driving the occurrence of claims, while accounting for the exposure in a specific occurrence period. The number of claims for occurrence period i, say K_i, follows a Poisson process with occurrence measure $\theta w(i)$. $w(i)$ is the exposure measure for occurrence period i ($i = 1, \ldots, I$). However, since we only observe reported claims, the Poisson process should be thinned in the following way

$$\theta w(i) F_1(t_i^* - 1; \boldsymbol{\nu}), \tag{1}$$

where t_i^* denotes the number of periods between the occurrence period i and the evaluation date. As introduced in Section 2.1.1, $F_1(.)$ is the cdf assumed for reporting delay.

2.1.3 Development Pattern

Structuring the development pattern Let the random variable Y_{ikj} represent the j^{th} incremental partial amount for the k^{th} claim ($k = 1, \ldots, K_i$) from occurrence period i ($i = 1, \ldots, I$). We obtain the total cumulative amount paid for claim (ik) by multiplying the initial amount, Y_{ik1}, by one or more *link ratios*. The initial amount, together with the vector of link ratio(s), forms the *development pattern* of the claim. This approach is similar to the one used in the chain–ladder model (see Mack (1993) and Mack (1999)).

However, with chain–ladder, the index j is for development period instead of partial payment. Using a *development–to–development* period model (as chain–ladder does) with individual claims can be problematic because the length of the development pattern differs among claims, and many development factors will have value one. We avoid this in the *payment–to–payment* approach used in our paper.

For a claim (ik) with a strict positive value of $U_{ik} = u_{ik}$, the vector $\mathbf{\Lambda}^{(ik)}_{u_{ik}+1}$ of length $u_{ik}+1$ gives the development pattern

$$\mathbf{\Lambda}^{(ik)}_{u_{ik}+1} = \begin{bmatrix} Y_{ik1} & \lambda^{(ik)}_1 & \dots & \lambda^{(ik)}_{u_{ik}} \end{bmatrix}', \tag{2}$$

where

$$\lambda^{(ik)}_j = \frac{\sum_{r=1}^{j+1} Y_{ikr}}{\sum_{r=1}^{j} Y_{ikr}}, \tag{3}$$

for $j = 1, \dots, u_{ik}$. In the stochastic version of the chain–ladder model, successive development factors are supposed to be non–correlated given past information. Moreover, independence is assumed between the initial payment and the vector of development factors. The so–called PIC model from Merz and Wüthrich (2010) is an exception. The study by Happ and Merz (2012) examines dependence structures for link ratios in the PIC model. In the individual framework developed in our paper assuming independence is problematic and unrealistic (as demonstrated empirically in Section 4.2.3, Figure 8 (Bodily Injury) and 11 (Material Damage)). This motivates the use of a flexible multivariate distribution for $\mathbf{\Lambda}^{(ik)}_{u_{ik}+1}$ ($i = 1, \dots, I$ and $k = 1, \dots, K_i$). Such a distribution should be able to account for the dependence present in the development pattern vectors, as well as the asymmetry in each of its components.

A flexible multivariate distribution for the development pattern. Our paper uses the family of Multivariate Skew Symmetric (MSS) distributions (see Gupta and Chen (2004) and Deniz (2009)) to model the development pattern of a claim (ik) on log scale. More specifically, we will use the Multivariate Skew Normal (MSN) distribution as multivariate versions of the Univariate Skew Normal (USN) distribution (from Roberts and Geisser (1966) and Azzalini (1985)).

Definition 2.1 (MSS and MSN distribution.) *Let $\boldsymbol{\mu} = [\mu_1 \ \dots \ \mu_k]'$ be a vector of location parameters, $\mathbf{\Sigma}$ a $(k \times k)$ positive definite symmetric scale matrix and $\mathbf{\Delta} = [\Delta_1 \ \dots \ \Delta_k]'$ a vector of shape parameters. The $(k \times 1)$ random vector $\mathbf{X}$ follows a Multivariate Skew*

Symmetric (MSS) distribution if its density function is of the form

$$MSS\left(\mathbf{X};\boldsymbol{\mu},\boldsymbol{\Sigma}^{1/2},\boldsymbol{\Delta}\right) = \frac{2^k}{det(\boldsymbol{\Sigma})^{1/2}} g^*\left(\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\mu})\right)\prod_{j=1}^{k} H\left(\Delta_j \mathbf{e}_j' \boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\mu})\right), \tag{4}$$

where $g^*(\mathbf{x}) = \prod_{j=1}^{k} g(x_j)$*,* $g(\cdot)$ *is a density function symmetric around* 0*,* $H(\cdot)$ *is an absolutely continuous cumulative distribution function with* $H'(\cdot)$ *symmetric around* 0 *and* $\mathbf{e}_j'$ *are the elementary vectors of the coordinate system* $\mathbb{R}^k$*.* [1]

The Multivariate Skew Normal (MSN) distribution is obtained from (4) *by replacing* $g(\cdot)$ *and* $H(\cdot)$ *with the pdf and cdf of the standard Normal distribution, respectively.*

2.2 The Likelihood

For the sake of clarity, the likelihood function will be divided into three parts: an expression for the likelihood of closed, RBNP and RBNS claims.

Closed claims. For closed claims (Cl), the likelihood function is given below. Hereby, t_{ik}^* refers to the evaluation date, expressed as number of periods after occurrence. $(ik)_{\text{Cl}}$ refers to a closed claim.

$$\begin{aligned}
\mathcal{L}^{\text{Cl}} \propto\ & \prod_{(ik)_{\text{Cl}}} \text{MSS}(\ln\left(\boldsymbol{\Lambda}_{u_{ik}+1}\right); \boldsymbol{\mu}_{u_{ik}+1}, \boldsymbol{\Sigma}_{u_{ik}+1}^{1/2}, \boldsymbol{\Delta}_{u_{ik}+1}|u_{ik}) \cdot f_1(t_{ik};\boldsymbol{\nu}|T_{ik} \leq t_{ik}^* - 1) \\
\cdot & \prod_{(ik)_{\text{Cl}}} f_2(q_{ik};\psi|Q_{ik} \leq t_{ik}^* - t_{ik} - 1) \cdot f_3(u_{ik};\boldsymbol{\beta}|U_{ik} \leq t_{ik}^* - q_{ik} - t_{ik} - 1) \\
\cdot & \prod_{(ik)_{\text{Cl}}} \{I(u_{ik}=0)(1)\} \cdot \{I(u_{ik} \geq 1) f_4(n_{ik1};\phi|0 < N_{ik1} \leq t_{ik}^* - t_{ik} - q_{ik} - u_{ik})\} \\
\cdot & \{I(u_{ik} \geq 2)\prod_{j=2}^{u_{ik}} f_4(n_{ikj};\phi|0 < N_{ikj} \leq t_{ik}^* - t_{ik} - q_{ik} - (u_{ik} - j + 1) - \sum_{p=1}^{j-1} n_{ikp}).\}
\end{aligned} \tag{5}$$

The first component in this likelihood (i.e. 'MSS(. . .)') is the multivariate distribution of the development pattern, given the total number of link ratio(s). The other components, $f_1(.)$, $f_2(.)$, $f_3(.)$ and $f_4(.)$, refer to reporting delay, first payment delay, the number of periods with payment and the delay between two periods with payment. The random variables involved in the time structure (T, Q, U and N) have their distribution censored at the evaluation date.

[1]The scale parameter $\boldsymbol{\Sigma}$ is not the usual variance-covariance matrix as in the Multivariate Normal distribution. A MSS random vector is defined by $\boldsymbol{\Sigma}^{1/2}$ in place of $\boldsymbol{\Sigma}$ because of the plurality of the square roots of $\boldsymbol{\Sigma}$. Without subscript, $\boldsymbol{\Sigma}^{1/2}$ designs any square root of the matrix $\boldsymbol{\Sigma}$.

RBNS claims. For Reported But Not Settled claims (RBNS), the likelihood is (with u_{ik}^* the observed number of periods with payment after the first one, and $(ik)_{\text{RBNS}}$ indicating an RBNS claim)

$$\begin{aligned}
\mathcal{L}^{\text{RBNS}} \quad \propto & \prod_{(ik)_{\text{RBNS}}} \text{MSS}(\ln\left(\boldsymbol{\Lambda}_{u_{ik}^*+1}\right); \boldsymbol{\mu}_{u_{ik}^*+1}\boldsymbol{\Sigma}^{1/2}_{u_{ik}^*+1}, \boldsymbol{\Delta}_{u_{ik}^*+1}|u_{ik}^*) \cdot f_1(t_{ik};\boldsymbol{\nu}|T_{ik} \leq t_{ik}^* - 1) \\
\cdot & \prod_{(ik)_{\text{RBNS}}} f_2(q_{ik};\psi|Q_{ik} \leq t_{ik}^* - t_{ik} - 1) \cdot (1 - F_3(u_{ik}^* - 1;\boldsymbol{\beta})) \\
\cdot & \prod_{(ik)_{\text{RBNS}}} I(u_{ik}^* = 0)(1) \cdot I(u_{ik}^* \geq 1) f_4(n_{ik1};\phi|0 < N_{ik1} \leq t_{ik}^* - t_{ik} - q_{ik} - u_{ik}^*) \\
\cdot & \{I(u_{ik}^* \geq 2) \prod_{j=2}^{u_{ik}^*} f_4(n_{ikj};\phi|0 < N_{ikj} \leq t_{ik}^* - t_{ik} - q_{ik} - (u_{ik}^* - j + 1) - \sum_{p=1}^{j-1} n_{ikp})\}.
\end{aligned} \tag{6}$$

RBNP claims. Finally, for Reported But Not Paid claims (RBNP), the likelihood function is (with $(ik)_{\text{RBNP}}$ indicating an RBNP claim)

$$\mathcal{L}^{\text{RBNP}} \quad \propto \quad \prod_{(ik)_{\text{RBNP}}} f_1(t_{ik};\boldsymbol{\nu}|T_{ik} \leq t_{ik}^* - 1) \cdot (1 - F_2(t_{ik}^* - t_{ik} - 1;\psi)). \tag{7}$$

2.3 Analytical Results for Best Estimates of Outstanding Reserves

The model specified in Section 2.1 and 2.2 allows to derive analytical results for the n^{th} moment of an IBNR, RBNP and RBNS claim, as well as for the expected value of the IBNR, RBNP and RBNS reserve. Proofs are deferred to Appendix A. We drop the (ik) subscript for reasons of simplicity.

Proposition 2.2 (n^{th} moment of an IBNR or RBNP claim.) *Let C be the random variable representing the total claim amount of an IBNR (or RBNP) claim*

$$C \quad = \quad Y_1 \cdot \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_U. \tag{8}$$

Using the model assumptions from Section 2.1 and 2.2 with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and shape vector $\boldsymbol{\Delta}$, the n^{th} moment of C is given by

$$E\left[2^{U+1}\exp\left(\mathbf{t}_n'\boldsymbol{\mu}_{U+1} + 0.5\mathbf{t}_n'\boldsymbol{\Sigma}^{1/2}_{U+1}\left(\boldsymbol{\Sigma}^{1/2}_{U+1}\right)'\mathbf{t}_n\right) \cdot \prod_{j=1}^{U+1}\Phi\left(\frac{\Delta_j \cdot \left(\left(\boldsymbol{\Sigma}^{1/2}_{U+1}\right)'\mathbf{t}_n\right)_j}{\sqrt{1+\Delta_j^2}}\right)\right]_U. \tag{9}$$

$\mathbf{t}_n$ is an $((U+1) \times 1)$ vector, specified as $[n\ n\ \ldots\ n]'$.

Proposition 2.3 gives the corresponding result for an RBNS claim. The distinguishing feature between Proposition 2.2 and 2.3 is the fact that for an RBNS claim part of the development pattern is already observed.

Proposition 2.3 (n^{th} moment of an RBNS claim.) *Define*

$$\begin{aligned}\boldsymbol{\Lambda}_{U+1} &= \begin{bmatrix}\boldsymbol{\Lambda}_A\\ \boldsymbol{\Lambda}_B\end{bmatrix}, \quad \boldsymbol{\mu}_{U+1} = \begin{bmatrix}\boldsymbol{\mu}_A\\ \boldsymbol{\mu}_B\end{bmatrix},\\ \boldsymbol{\Sigma}_{U+1}^{1/2} &= \begin{bmatrix}\boldsymbol{\Sigma}_{AA} & \mathbf{0}\\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB}\end{bmatrix}, \quad \boldsymbol{\Delta}_{U+1} = \begin{bmatrix}\boldsymbol{\Delta}_A\\ \boldsymbol{\Delta}_B\end{bmatrix},\end{aligned} \tag{10}$$

where $\boldsymbol{\Lambda}_A$, $\boldsymbol{\mu}_A$ and $\boldsymbol{\Delta}_A$ are $U_A \times 1$ (with $U_A < U+1$). $\boldsymbol{\Sigma}_{AA}$ is a $U_A \times U_A$ lower triangular matrix with positive diagonal elements and $\boldsymbol{\Sigma}_{BB}$ is a $U_B \times U_B$ lower triangular matrix with positive diagonal elements. Hereby, $U_B + U_A = U+1$, the total number of periods with partial payment.

We define $[\boldsymbol{\mu}^*_{U+1}|\boldsymbol{\Lambda}_A = \boldsymbol{\ell}_A] := \boldsymbol{\mu}_B + \boldsymbol{\Sigma}_{BA}\boldsymbol{\Sigma}_{AA}^{-1}(\boldsymbol{\ell}_A - \boldsymbol{\mu}_A)$, $\boldsymbol{\Sigma}^*_{U+1} = \boldsymbol{\Sigma}_{BB}$ and $\boldsymbol{\Delta}^*_{U+1} = \boldsymbol{\Delta}_B$. The conditional final amount of a claim C, given past information, is defined as

$$[C|\boldsymbol{\Lambda}_A = \boldsymbol{\ell}_A] = y_1 \cdot \ell_1 \cdot \ldots \cdot \ell_{u_A - 1} \cdot \lambda_{u_A} \cdot \ldots \cdot \lambda_U. \tag{11}$$

Using the model assumptions from Section 2.1 and 2.2, the n^{th} moment of C is given by

$$\begin{aligned}&E[C^n|\boldsymbol{\Lambda}_A = \boldsymbol{\ell}_A] = \left(y_1 \cdot \ell_1 \cdot \ldots \cdot \ell_{u_A-1}\right)^n\\ &\cdot \; E\left[2^{U_B}\exp\left(\boldsymbol{h}_n'\boldsymbol{\mu}^*_{U+1} + 0.5\boldsymbol{h}_n'\left(\boldsymbol{\Sigma}^*_{U+1}\right)^{1/2}\left(\left(\boldsymbol{\Sigma}^*_{U+1}\right)^{1/2}\right)'\boldsymbol{h}_n\right) \cdot \prod_{j=1}^{U_B}\Phi\left(\frac{\Delta_j^* \cdot \left(\left(\left(\boldsymbol{\Sigma}^*_{U+1}\right)^{1/2}\right)'\boldsymbol{h}_n\right)_j}{\sqrt{1+\left(\Delta_j^*\right)^2}}\right)\right]_{U_B}\end{aligned} \tag{12}$$

with the $(U_B \times 1)$ vector $\boldsymbol{h}_n := [n\ n\ \ldots\ n]'$.

Analytical expressions for the total outstanding IBNR, RBNP and RBNS reserve follow immediately from Proposition 2.2 and 2.3.

Proposition 2.4 (Best estimates for the IBNR, RBNP and RBNP reserves.) *Let $\mathcal{I}$ denote the observed information for all claims in the data set. We define $\mathbf{t}_n$, $\mathbf{h}_n$, $\boldsymbol{\mu}^*_{U+1}$, $\boldsymbol{\Sigma}^*_{U+1}$ and $\boldsymbol{\Delta}^*_{U+1}$ as in Proposition 2.2 and 2.3, respectively. Using the model assumptions from Section 2.1 and 2.2, the best estimate of the outstanding IBNR, RBNP and RBNS reserves follow.*

(a) The expected value of the total amount outstanding for IBNR and RBNP claims,

respectively, is

$$
\begin{aligned}
& E[IBNR|\mathcal{I}] \;\; \textit{versus}\; E[RBNP|\mathcal{I}] \\
= \;\; & (x) \cdot E\left[2^{U+1} \exp(\mathbf{t}_1' \boldsymbol{\mu}_{U+1} + 0.5\mathbf{t}_1' \boldsymbol{\Sigma}_{U+1}^{1/2} \left(\boldsymbol{\Sigma}_{U+1}^{1/2}\right)' \mathbf{t}_1) \cdot \prod_{j=1}^{U+1} \Phi\left(\frac{\Delta_j \cdot ((\boldsymbol{\Sigma}_{U+1}^{1/2})' \mathbf{t}_1)_j}{\sqrt{1+\Delta_j^2}}\right)\right]_U,
\end{aligned}
\tag{13}
$$

where (x) should be replaced with $E[K_{IBNR}]$ in case of IBNR reserves, and with k_{RBNP}, the observed number of open claims without payment, in case of RBNP reserves. The expected number of IBNR claims follows from the Poisson process driving the occurrence of claims (appropriately thinned to represent IBNR claims).

(b) *The expected value of the total amount outstanding for RBNS claims is*

$$
\begin{aligned}
& E[RBNS|\mathcal{I}] \\
= \;\; & \sum_{(ik)_{RBNS}} y_1 \cdot \ell_1 \cdot \ldots \cdot \ell_{u_1-1} \\
\cdot \;\; & E\left[2^{U_B} \exp(\mathbf{h}_1' \boldsymbol{\mu}_{U+1}^* + 0.5\mathbf{h}_1' \left(\boldsymbol{\Sigma}_{U+1}^*\right)^{1/2} \left(\left(\boldsymbol{\Sigma}_{U+1}^*\right)^{1/2}\right)' \mathbf{h}_1) \cdot \prod_{j=1}^{U_B} \Phi\left(\frac{\Delta_j^* \cdot \left(\left(\left(\boldsymbol{\Sigma}_{U+1}^*\right)^{1/2}\right)' \mathbf{h}_1\right)_j}{\sqrt{1+\left(\Delta_j^*\right)^2}}\right)\right]_{U_B},
\end{aligned}
\tag{14}
$$

where the sum goes over all RBNS claims.

3 The Data

3.1 Background

We study the data set from Antonio and Plat (2012) on a portfolio of general liability insurance policies for private individuals [2]. Available information is from January 1997 till December 2004. Originally, information is available till August 2009, but to enable out–of–sample prediction we remove the observations from January 2005 to August 2009. Two types of payments are registered in the data set: Bodily Injury (BI) and Material Damage (MD) [3]. Figure 2 represents the development of a random claim from the data set. Following the approach presented in this paper, Figure 3 transforms the data set to discrete time periods (here: one period is one year).

[2]As in Antonio and Plat (2012) we discount payments to 1/1/1997 with the appropriate consumer price index.

[3]In contrast with Antonio and Plat (2012) a claim can have both BI payments, as well as MD payments. In Antonio and Plat (2012) a claim with at least one BI payment was considered as BI.

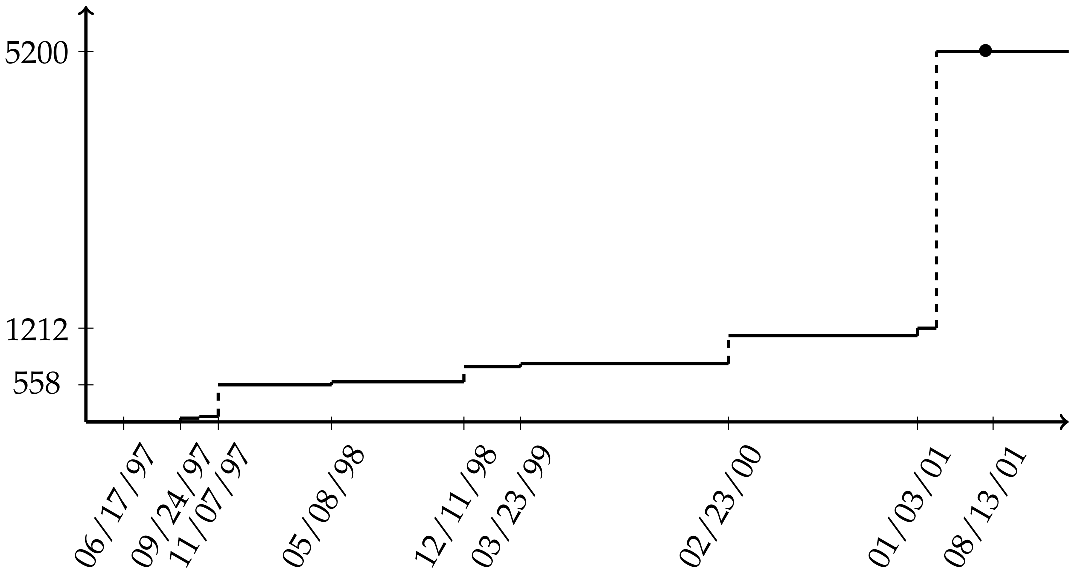


Figure 2: Development of a random claim in continuous time. The x–axis represents the date of each event and the y–axis represents the cumulative amount paid for the claim.

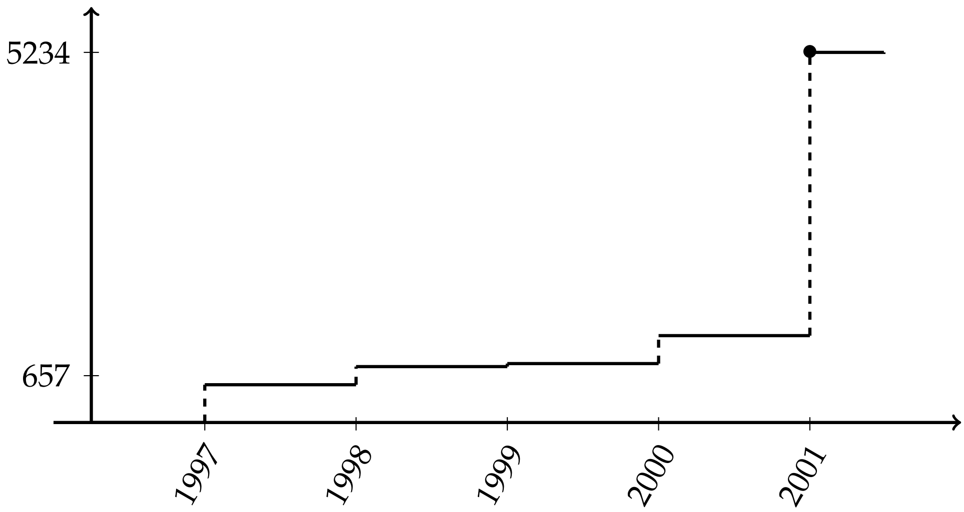


Figure 3: Development of the claim from Figure 2 in a discrete time framework (annual).

The accident occurs at 06/17/1997. The claim is reported to the company on 07/22/1997, thus: $t_{(ik)} = 0$. A first payment is done on 09/24/1997, implying a first payment delay of 0 periods ($q_{(ik)} = 0$). Consequently, payments follow on 10/21/1997, 11/07/1997, 05/08/1998, 12/11/1998, 03/23/1999, 02/23/2000, 01/03/2001 and 02/24/2001. Therefore, $u_{(ik)} = 4$ and $n_{(ik),1} = n_{(ik),2} = n_{(ik),3} = n_{(ik),4} = 1$. Closure is at 08/13/2001, thus

$n_{(ik),5} = 0$.

3.2 Descriptive Statistics

The data set consists of 279,094 claims; 273,977 claims are related to Material Damage (MD) and 5,117 claims to Bodily Injury (BI). 268,484 MD claims (181,828 with at least one payment and 86,656 with no payment) and 4,098 BI claims (2,961 with at least one payment and 1,137 with no payment) are closed in the data set. We present descriptive statistics for closed claims with positive payments in Table 1. In Section 4.1 descriptive graphics follow representing reporting delay, first payment delay and the number of periods with payment (see Figure 5). We illustrate correlation between development factors in Figures 8 (Bodily Injury) and 11 (Material Damage).

Class	Variables	Mean	Median	s.e.	Minimum	Maximum	Number of Observations
BI	Y_1	1,008	351	3,274	0.18	148,900	2,961
	λ_1	10.24	3.23	31.52	1.01	653.33	991
	λ_2	4.50	1.95	10.80	1.00	127.74	253
	λ_3	2.73	1.80	2.18	1.00	11.94	89
	λ_4	2.67	1.92	2.22	1.00	11.44	37
	Total Claim	2,961	624	11,825	6.3	410,500	2,961
MD	Y_1	298	151	528	0.35	68,810	181,828
	λ_1	5.44	2.18	11.71	1.00	371.40	1,555
	λ_2	2.16	1.41	1.73	1.01	6.93	13
	Total Claim	305	153	679	0.35	108,300	181,828

Table 1: Descriptive statistics for closed claims: first payment, total claim amount, and development factors λ_i with $i \leq 4$ for BI claims and $i \leq 2$ for MD claims.

4 Distributional Assumptions and Estimation Results

4.1 Distributional Assumptions

Distributions for number of periods For the random variables describing the time structure part of a claim's development (i.e. $\{T_{ik}\}$, $\{Q_{ik}\}$, $\{U_{ik}\}$ and $\{N_{ik}\}$ from Section 2.1), we consider mixtures of a discrete distribution with degenerate components (similar to Antonio and Plat (2012)). For reporting delay, for instance, we investigate distributions of the following type

$$f_1(t;\boldsymbol{\nu}) = \sum_{s=0}^{p} \nu_s I_s(t) + \left(1 - \sum_{s=0}^{p} \nu_s\right) f_{T|T>p}(t), \tag{15}$$

where $I_s(t) = 1$ for reporting in the s^{th} period after the period of occurrence and 0 otherwise. $f(t)$ is the pdf of a discrete distribution with parameter(s) $\nu_{p+1}, \ldots, \nu_{p+q}$.

Further on, we investigate the use of a Geometric, Binomial, Poisson and Negative Binomial distribution for $f(.)$, combined with different values for p ($p = 0, 1, 2, 3$).

Development pattern For the logarithm of the development pattern vector (as in (2)), we consider the MSN distribution on the one hand and the special case where $\mathbf{\Delta} = \mathbf{0}$, i.e. the Multivariate Normal distribution (MN), on the other hand. The following structures are considered for the $z \times z$ matrix $\Sigma_c^{1/2}$ [4]: unstructured (UN), Toeplitz (TOEP), Compound Symmetry (CS) and Diagonal (DIA) (see below).

$$\underset{\text{(UN)}}{\begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ \sigma_{21} & \sigma_2^2 & 0 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ \vdots & & & \ddots & \vdots \\ \sigma_{z1} & \sigma_{z2} & \sigma_{z3} & \dots & \sigma_z^2 \end{pmatrix}} \qquad \underset{\text{(TOEP)}}{\begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & 0 & \dots & 0 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3^2 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ \sigma_z\sigma_1\rho_{z-1} & \sigma_z\sigma_2\rho_{z-2} & \sigma_z\sigma_3\rho_{z-3} & \dots & \sigma_z^2 \end{pmatrix}}$$

$$\underset{\text{(CS)}}{\begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & 0 & \dots & 0 \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ \sigma_z\sigma_1\rho & \sigma_z\sigma_2\rho & \sigma_z\sigma_3\rho & \dots & \sigma_z^2 \end{pmatrix}} \qquad \underset{\text{(DIAG)}}{\begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^2 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_z^2 \end{pmatrix}}$$

4.2 Estimation Results

Following the discussion and approach in Antonio and Plat (2012), we fit the model separately for Material Damage and Bodily Injury payments. We perform data manipulations and likelihood optimization with R (using additional packages, like ChainLadder and sn for the Skew Normal distribution). We use numerical approximations of the Hessian matrix to estimate standard errors. For each component in the model, a model selection step is performed, comparing different models based on AIC and BIC. We highlight selected model specifications in blue in the tables following below.

4.2.1 Distributions for number of periods

For the discrete random variables $\{T_{ik}\}$, $\{Q_{ik}\}$, $\{U_{ik}\}$ and $\{N_{ik}\}$, we investigate the use of a mixture of p degenerate distributions with a basic count distribution (see (15)). Consequently, $p + q + 1$ parameters have to be estimated for each variable. Our model

[4] For MSN and MN, matrix $\Sigma_c^{1/2}$ refers to the square root of the covariance matrix $\mathbf{\Sigma}$, as obtained by Cholesky decomposition.

selection procedure (based on AIC and BIC) prefers the use of a Geometric distribution, combined with degenerate components. Figure 6 shows model selection steps assisting in the choice of the number of degenerate components. Figure 7 displays parameter estimates and standard errors for the preferred specifications. Observed and estimated results are compared in Figure 5, at least for the components necessary to project claims till settlement.

4.2.2 Occurrence of claims

Using the distributions selected for reporting delay, we estimate the thinned Poisson process from (1). Hereby, the exposure measure $w(.)$ is expressed in years. Results are: $\hat{\theta}_{BI} = 0.7445$ (s.e. 0.02) and $\hat{\theta}_{MD} = 38.96$ (s.e. 0.11).

4.2.3 Development pattern

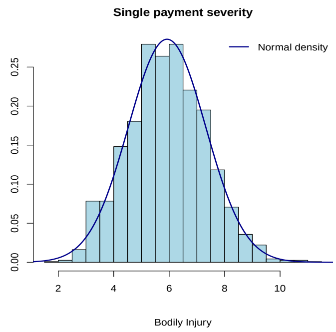
The development consists of a single payment. For the logarithm of the severity of the first and only payment, we explore the use of a Univariate Skew Normal (USN) as well as a Normal (N) distribution. The estimation results and a graphical goodness–of–fit check are in Figure 4. For the data at hand, the Normal distribution is to be preferred.

The development consists of more than one payment. We examine the use of the Multivariate Skew Normal (MSN), as well as the Multivariate Normal (MN), distribution for the logarithm of the development pattern vector $\mathbf{\Lambda}^{(ik)}_{U_{ik}+1}$ (see (2)). For Bodily Injury we restrict the maximal dimension of the development vector, say m_p, to 5 and to $m_p = 3$ for Material Damage [5]. Therefore, we fit a location vector of dimension $m_p \times 1$, a scale matrix of dimension $m_p \times m_p$ and a shape vector of dimension $m_p \times 1$. When observed claims use less development factors, appropriate subvectors and submatrices are used in the likelihood. If the simulated number of periods with payment is bigger than m_p, we apply a tail factor [6]. Figures 9 and 10 (Bodily Injury) and 12 and 13 (Material Damage) present results of the model selection steps, as well as parameter estimates for the preferred Multivariate Skew Normal and the preferred Multivariate Normal distribution. Empirical data and contour plots for the chosen MSN multivariate density are compared in Figure 8 (Bodily Injury) and 11 (Material Damage).

[5] In the data set we observe only 8 BI claims with more than 5 periods with payment and 2 MD claims with more than 3 periods with payment.

[6] This tail factor is the geometric average of empirically observed development factors.

	BI		MD	
	USN	N	USN	N
μ	5.9377	5.9226	4.9541	5.0428
(s.e.)	(1.04)	(0.03)	(0.06)	($<$ 0.01)
σ	1.3966	1.3968	1.1663	1.1637
(s.e.)	(0.02)	(0.02)	(0.01)	($<$ 0.01)
δ	-0.0139	-	0.0959	-
(s.e.)	(0.94)	-	(0.07)	-
AIC	4,124	4,122	284,855	284,853
BIC	4,141	4,133	284,885	284,873



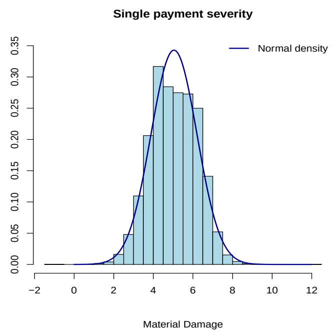


Figure 4: Logarithm of the severity of the first and only payment: (on the left) estimation results for the Univariate Skew Normal distribution (USN) (with parameters μ, σ and scale parameter δ) and the Normal distribution (N) (with parameters μ and σ); (on the right) empirical and fitted densities for Bodily Injury (top) and Material Damage (bottom).

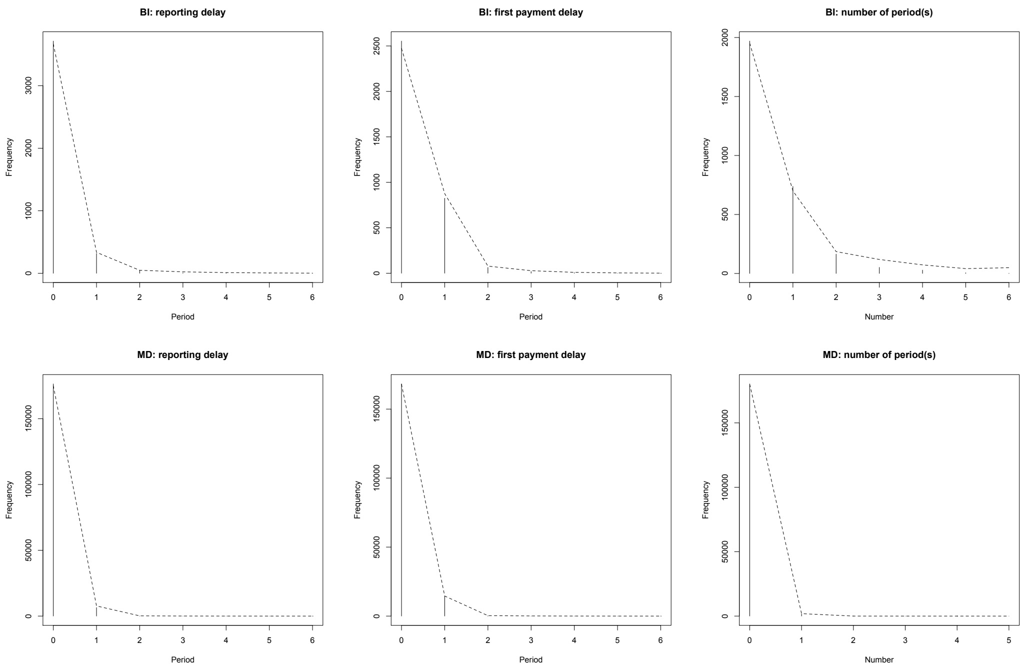


Figure 5: Observed and estimated frequency distributions for Bodily Injury (BI, top row) and Material Damage (MD, bottom row). From left to right: reporting delay, first payment delay and number of intermediate payments after the first one.

		BI		MD	
	p	AIC	BIC	AIC	BIC
$(T;\nu)$	baseline	3,120	3,126	88,730	88,740
	0	2,987	3,000	85,714	85,735
	1	2,966	2,985	85,484	85,515
	2	2,961	2,987	85,479	85,521
	3	2,963	2,994	85,485	85,537
$(Q;\psi)$	baseline	4,882	4,888	116,611	116,621
	0	4,605	4,617	111,045	111,066
	1	4,575	4,594	110,680	110,711
	2	4,577	4,602	110,676	110,717
	3	4,578	4,609	110,680	110,732
$(U;\beta)$	baseline	6,102	6,108	18,255	18,265
	0	6,096	6,108	18,250	18,270
	1	6,025	6,043	18,233	18,264
	2	6,026	6,051	18,233	18,273
	3	6,017	6,048	18,235	18,285

Figure 6: Model selection for $\{T_{ik}\}$, $\{Q_{ik}\}$, $\{U_{ik}\}$, using the structure from (15) with a Geometric distribution for the basic count distribution.

Class	Parameter Index	Report delay $(T;\nu_s)$ (s.e.)	First pmt delay $(Q;\psi_s)$ (s.e.)	Number partial pmt $(U;\beta_s)$ (s.e.)
	0	0.8953	0.7127	0.5192
		(< 0.001)	(< 0.001)	(0.010)
	1	0.0819	0.2522	0.2470
		(0.003)	(0.003)	(0.008)
BI	2	0.5144	0.6431	0.3094
		(0.064)	(0.052)	(0.022)
	0	0.9565	0.9181	0.9896
		(< 0.001)	(< 0.001)	(< 0.001)
	1	0.0421	0.0794	0.0103
		(< 0.001)	(< 0.001)	(< 0.001)
MD	2	0.6820	0.6729	0.7184
		(0.031)	(0.026)	(0.125)

Figure 7: Estimation results for the selected distribution for $\{T_{ik}\}$, $\{Q_{ik}\}$, $\{U_{ik}\}$, i.e. a Geometric distribution with degenerate components. Parameters are denoted as in (15).

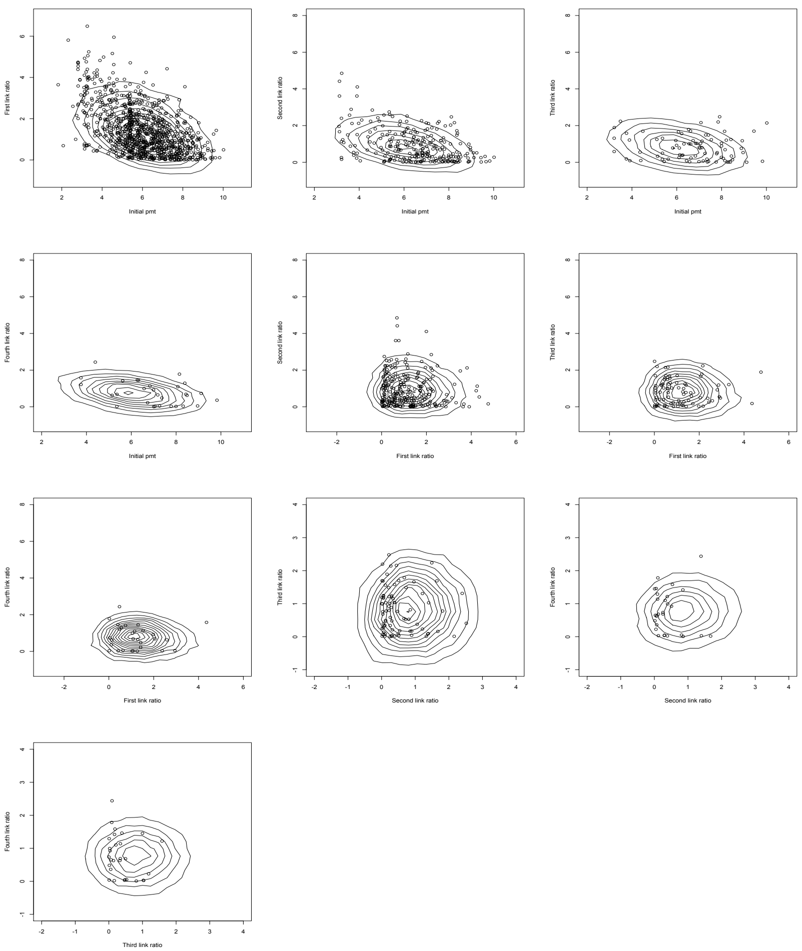


Figure 8: Bodily injury: empirical observations of the development vector (2) and contour plots obtained from selected MSN model (see right). First row of plots (from left to right): first link ratio vs. initial payment, second link ratio vs. initial payment, third link ratio vs. initial payment. Second row (from left to right): fourth link ratio vs. initial payment, second vs. first link ratio, third vs. first link ratio. Third row (from left to right): fourth vs. first link ratio, third vs. second link ratio, fourth vs. second link ratio. Fourth row: fourth vs. third link ratio.

Model		# Parms.	$-ll$	AIC	BIC
MSN	UN	20	$3,431$	$6,902$	$7,000$
	TOEP	14	$3,435$	$6,897$	$6,966$
	CS	11	$3,444$	$6,910$	$6,964$
	DIA	10	$3,605$	$7,230$	$7,279$
MN	UN	20	$3,496$	$7,032$	$7,128$
	TOEP	14	$3,499$	$7,025$	$7,094$
	CS	11	$3,531$	$7,083$	$7,137$
	DIA	10	$3,723$	$7,465$	$7,514$

Figure 9: Bodily Injury: model selection steps examining MSN and MN specifications for the development pattern vector.

MSN Model			MN Model	
Location $\boldsymbol{\mu}$ (s.e.)	Scale $\boldsymbol{\Sigma}_c^{1/2}$	Shape $\boldsymbol{\Delta}$	Location $\boldsymbol{\mu}$ (s.e)	Scale $\boldsymbol{\Sigma}_c^{1/2}$
$\mu_1 = 5.44$	$\sigma_1 = 1.27$	$\Delta_1 = 0.51$	$\mu_1 = 6.04$	$\sigma_1 = 1.23$
(0.05)	$\sigma_2 = 1.18$	$\Delta_2 = 2.64$	(0.05)	$\sigma_2 = 0.97$
$\mu_2 = 0.53$	$\sigma_3 = 1.00$	$\Delta_3 = 2.29$	$\mu_2 = 1.43$	$\sigma_3 = 0.86$
(0.03)	$\sigma_4 = 0.83$	$\Delta_4 = -0.32$	(0.04)	$\sigma_4 = 0.82$
$\mu_3 = 0.63$	$\sigma_5 = 0.69$	$\Delta_5 = -0.002$	$\mu_3 = 0.95$	$\sigma_5 = 0.69$
(0.05)	$\rho = -0.28$		(0.05)	$\rho_1 = -0.49$
$\mu_4 = 1.49$			$\mu_4 = 0.64$	$\rho_2 = -0.23$
(0.09)			(0.08)	$\rho_3 = -0.003$
$\mu_5 = 1.12$			$\mu_5 = 0.66$	$\rho_4 = -0.26$
(0.10)			(0.11)	

Figure 10: Bodily Injury: parameter estimates for preferred MSN and MN distributions.

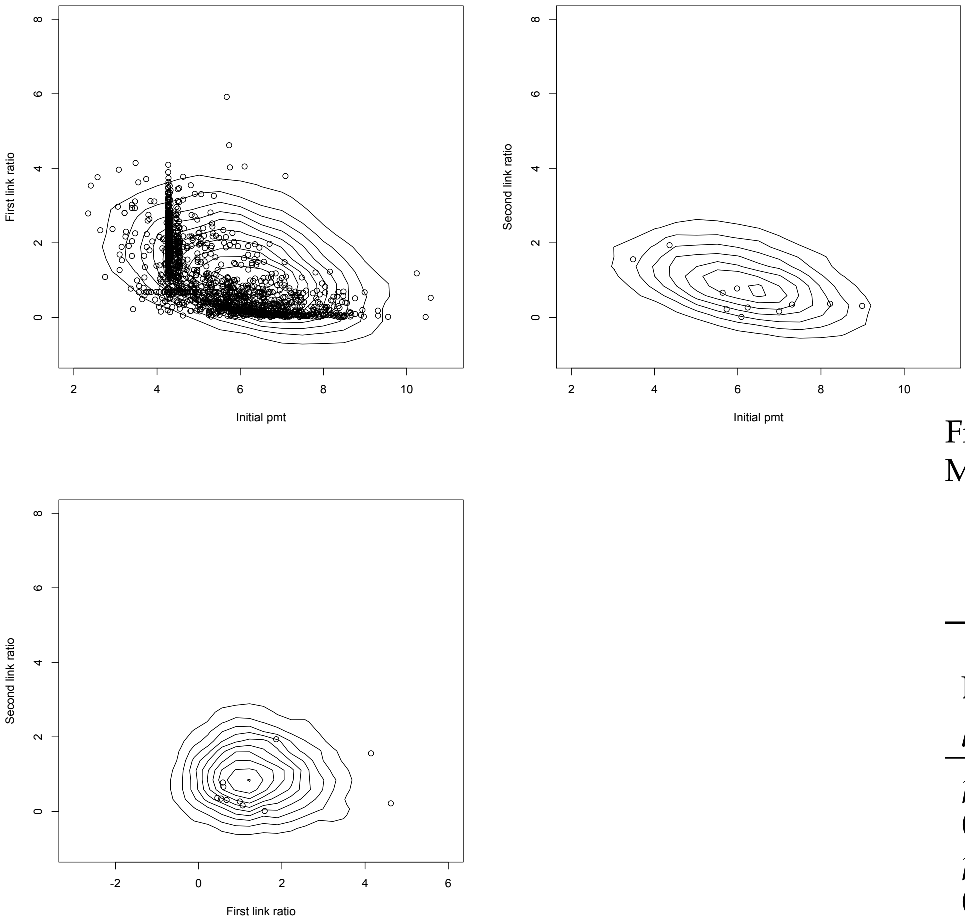


Figure 11: Material Damage: empirical observations of the development vector (2) and contour plots obtained from selected MSN model (see right). First row of plots (from left to right): first link ratio vs. initial payment, second link ratio vs. initial payment. Second row: second vs. first link ratio.

Model		# Parms.	$-ll$	AIC	BIC
MSN	UN	9	4,260	8,538	8,586
	TOEP	8	4,282	8,580	8,622
	CS	7	4,508	9,031	9,068
	DIA	6	4,740	9,492	9,524
MN	UN	9	4,260	8,538	8,586
	TOEP	8	4,271	8,557	8,600
	CS	7	4,510	9,033	9,071
	DIA	6	4,743	9,498	9,530

Figure 12: Material Damage: model selection steps examining MSN and MN specifications for the development pattern vector.

MSN Model			MN Model	
Location $\boldsymbol{\mu}$ (s.e.)	Scale $\boldsymbol{\Sigma}_c^{1/2}$	Shape $\boldsymbol{\Delta}$	Location $\boldsymbol{\mu}$ (s.e)	Scale $\boldsymbol{\Sigma}_c^{1/2}$
$\mu_1 = 5.44$	$\sigma_{11} = 1.27$	$\Delta_1 = -0.01$	$\mu_1 = 5.43$	$\sigma_{11} = 1.27$
(0.03)	$\sigma_{22} = 0.71$	$\Delta_2 = -0.01$	(0.03)	$\sigma_{22} = 0.71$
$\mu_2 = 1.12$	$\sigma_{33} = 0.75$	$\Delta_3 = 23.61$	$\mu_2 = 1.13$	$\sigma_{33} = 0.40$
(0.02)	$\sigma_{12} = -0.66$		(0.02)	$\sigma_{12} = -0.66$
$\mu_3 = 0.18$	$\sigma_{13} = -0.26$		$\mu_3 = 0.93$	$\sigma_{13} = -0.36$
(0.20)	$\sigma_{23} = -0.05$		(0.17)	$\sigma_{23} = -0.07$

Figure 13: Material Damage: parameter estimates for preferred MSN and MN distributions.

5 Prediction Results

We summarize the data set by occurrence and development year in run–off triangles, see Tables 2 and 3. Information with respect to occurrence years 2005 to 2009 (August) is available but not used in the analysis to enable out–of–sample prediction. This information is printed in bold in the run–off triangles.

Arrival	Development year							
year	1	2	3	4	5	6	7	8
1997	261	614	359	526	546	137	130	339
1998	202	473	307	336	269	56	179	**78**
1999	238	569	393	270	249	286	**132**	**97**
2000	237	557	429	496	406	**365**	**247**	**275**
2001	389	628	529	559	**446**	**375**	**147**	**239**
2002	260	570	533	**444**	**132**	**122**	**332**	**1,082**
2003	236	743	**558**	**237**	**217**	**205**	**171**	
2004	248	**794**	**401**	**236**	**254**	**98**		

Table 2: Incremental run-off triangle for Bodily Injury (in thousands).

Arrival	Development year							
year	1	2	3	4	5	6	7	8
1997	4,427	992	89	13	39	27	37	11
1998	4,389	984	60	35	76	24	0.5	**16**
1999	5,280	1,239	76	110	113	12	**0.4**	**0**
2000	5,445	1,164	172	16	6	**10**	**0**	**10**
2001	5,612	1,838	156	127	**13**	**3**	**0.4**	**3**
2002	6,593	1,592	74	**71**	**17**	**15**	**9**	**9**
2003	6,603	1,660	**150**	**52**	**37**	**18**	**3**	
2004	7,195	**1,417**	**109**	**86**	**39**	**15**		

Table 3: Incremental run-off triangle for Material Damage (in thousands).

5.1 Prediction of the IBNR and RBNP reserves

Best estimate for outstanding IBNR and RBNP reserves. Analytical expressions for the IBNR and RBNP reserve are available from Section 2.3, see Proposition 2.2, where unknown parameters should be replaced by estimates (as obtained in Section 4.2). Note that these expressions evaluate claims till settlement, even if this takes place beyond the boundary of the triangle. Table 5 displays these analytical results for Bodily Injury and Table 6 for Material Damage.

Simulation of outstanding IBNR and RBNP reserves. For each occurrence period, we simulate the number of IBNR claims (for Bodily Injury and Material Damage seperately) from a Poisson distribution with occurrence measure

$$\hat{\theta} w(i)(1 - F_1(t_i^* - 1; \hat{\boldsymbol{\nu}})). \tag{16}$$

Consequently, for each IBNR claim (denoted with (ik)), we simulate the number of period(s) with partial payments U_{ik} and the corresponding development pattern vector $\boldsymbol{\Lambda}_{U_{ik}}^{(ik)}$. Note that - with this strategy - we develop a claim till settlement (which can be beyond the boundary of the triangle). Taking the timing of partial payments into account would require simulation of the random variables T_{ik}, Q_{ik} and N_{ikj} (see Table 5 and 6 for results simulated until the boundary of the triangle).

The prediction routine for the RBNP reserve is similar to the routine for IBNR claims. However, the number of RBNP claims is observed, and therefore does not require a simulation step. The variable Q_{ik} should be simulated from a truncated distribution, using the condition $Q_{ik} > t_{ik}^* - t_{ik} - 1$.

Graphical results based on 5,000 simulations are shown in Figure 14. Tables 5 (Bodily Injury) and 6 (Material Damage) display corresponding numerical results.

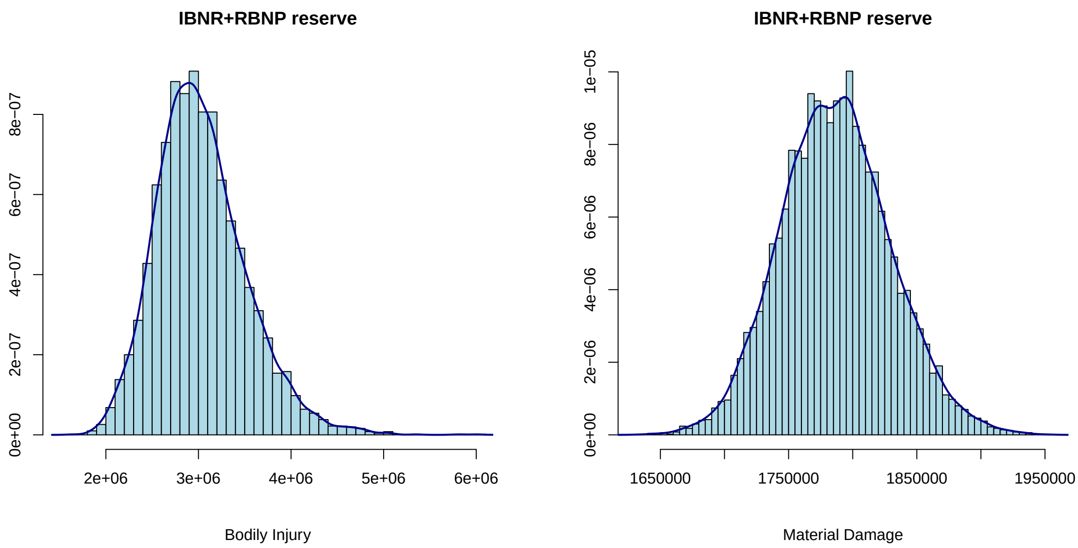


Figure 14: Histograms of the reserve obtained for IBNR and RBNP claims with the individual model for Bodily Injury (left) and Material Damage (right).

5.2 Prediction of the RBNS reserve

Best estimate for outstanding RBNS reserve. Tables 5 (Bodily Injury) and 6 (Material Damage) display analytical results for Bodily Injury and Material Damage payments.

Similar considerations apply as for IBNR and RBNP reserves.

Simulation of outstanding RBNS reserve. For each RBNS claim in the data set, we first simulate the number of period(s) with payment from the conditional pdf $f_3(u|u \geq u^*)$ where u^* is the observed number of periods with payment after the first one. Then, we simulate the missing part of the development pattern vector from the conditional MSN distribution (by conditioning on the observed part of the development pattern vector). Finally, we evaluate the RBNS reserve. Numerical results based on 5,000 simulations are in Tables 5 (Bodily Injury) and 6 (Material Damage) and corresponding graphical results are in Figure 15.

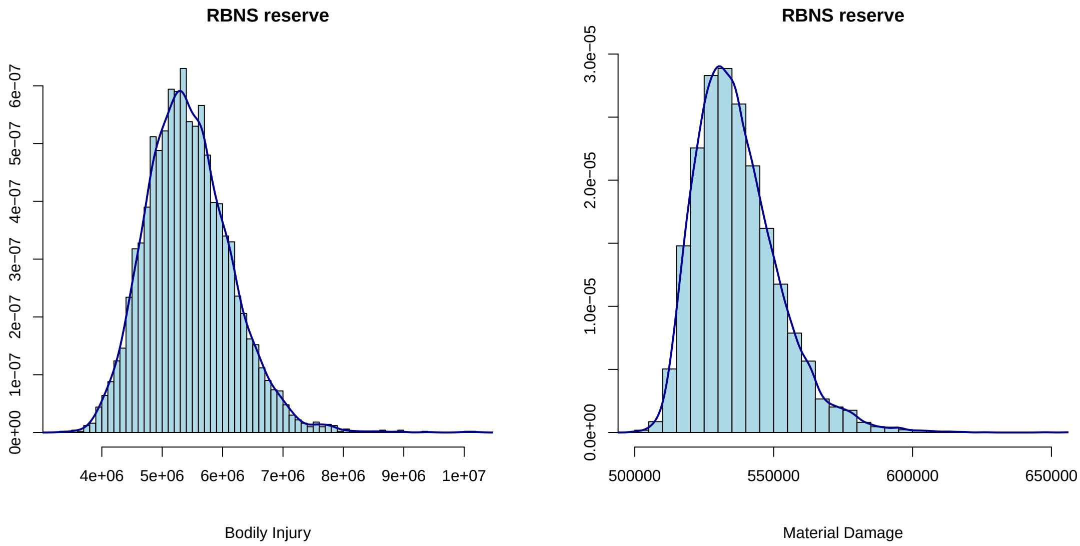


Figure 15: Histograms of the reserve obtained for RBNS claims with the individual model for Bodily Injury (left) and Material Damage (right).

5.3 Comparison of results

Tables 5 (for Bodily Injury) and 6 (for Material Damage) show prediction results obtained with our individual claims reserving method, as well as Mack's chain–ladder technique. The first two 'scenarios' in these tables display IBNR+RBNP (=IBNR$^+$), RBNS and Total reserves obtained with our preferred distributional assumptions (see Section 4.2.3) when claims are developed *until settlement*. Both analytical (first block of rows) and simulation based results (second block of rows) are given. The best estimate results obtained analytically are close to the mean of the corresponding predictive distribution obtained from simulation. This underpins the usefulness and appropriateness of the analytical formulas. The third block of rows shows simulation based results, including parameter uncertainty. This means that uncertainty in the location parameter was taken into account. The fourth block of rows gives simulation based results from the individual reserving method taking the policy limit of 2.5 MEuro into account (see Antonio and Plat (2012)). In a sixth block of rows we include simulation based results, accounting for policy limits, but restricting the development of claims to the right boundary of the triangle (i.e. development year 8), instead of developing claims until settlement. These results can be compared with Mack's chain–ladder results, as well as with the realized outcomes, displayed in bold in the lower triangles in Tables 2 and 3. Figure 17 illustrates this comparison. Results obtained with the chain–ladder method are represented by a lognormal density with mean and standard deviation as obtained from Mack's chain–ladder.

According to Figure 16 (simulation based, for Bodily Injury) and Table 4 (best estimate analytical results), the structure implied to $\boldsymbol{\Sigma}_c^{1/2}$ has minor impact on the resulting predictive distribution (obtained with MSN or MN assumption for (2)). However, the assumption of a Multivariate Normal versus Multivariate Skew Normal distribution for (2) has a clear impact on the predictive distribution of the outstanding reserves, at least for Bodily Injury payments. The impact is negligible for Material Damage (see Table 4). Recall from Figure 10 and 13 that all information criteria prefer the MSN distribution above the MN distribution. This sensitivity is a topic for future research.

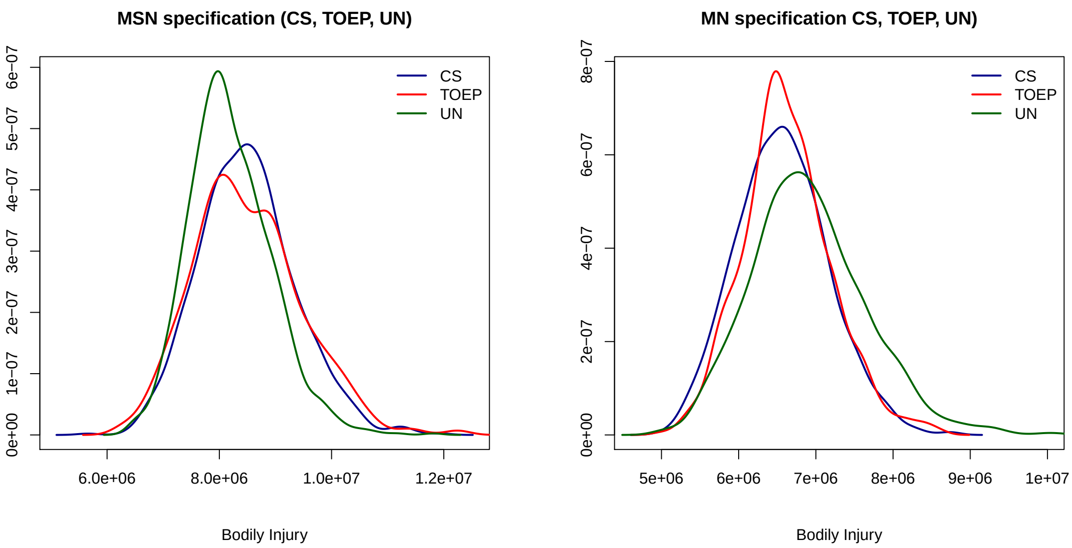


Figure 16: Bodily Injury: sensitivity of simulated predictive distributions with respect to the specification of the multivariate distribution for the development pattern vector. 'MSN' refers to Multivariate Skew Normal and 'MN' to Multivariate Normal.

The best estimate results reported in Tables 5 and 6 (simulation based, until the boundary of the triangle and taking the policy limit into account) are close to the results obtained in Antonio and Plat (2012). Our out–of–sample test (see Figure 17) demon-

MSN	UN	BI	8,132,051	MD	2,320,735
	TOEP		8,476,498		2,331,575
	CS		8,404,192		2,339,406
MN	UN	BI	6,836,694	MD	2,327,497
	TOEP		6,578,931		2,327,733
	CS		6,547,580		2,345,500

Table 4: Sensitivity of analytical best estimate results with respect to the specification of the multivariate distribution for the development pattern vector. 'MSN' refers to Multivariate Skew Normal and 'MN' to Multivariate Normal.

strates the usefulness of the method developed in this paper. As discussed in Antonio and Plat (2012), the lower triangle for Bodily Injury (see Table 2) shows an extreme payment (779,383 euro) in occurrence year 2002, development year 8. This is reflected in a realistic way by the individual loss reserving model.

6 Conclusions

This paper proposes a discrete time individual reserving model inspired by the chain–ladder model. The model is designed for a micro–level data set with the development of individual claims. Highlights of our approach are twofold. Firstly, on a claim by claim as well as aggregate level, analytical expressions for the first moment of the outstanding reserve are available. Secondly, the predictive distribution of the outstanding reserve is available by simulation. The latter approach allows to take policy characteristics, such as a policy limit, into account. The case study performed on a real–life general liability insurance portfolio demonstrates the usefulness of the model.

Several directions for future research can be envisaged. We plan further research with respect to the modeling of the first payment, using the Lognormal-Pareto distribution (see Pigeon and Denuit (2011)). Further investigation of the multivariate distribution for the development pattern vector is necessary. Nonparametric density estimation, as well as a copula approach, may be useful here. More precise modeling of inflation effects and inclusion of the 'time value of money' will be of importance in future work. Studying the approach in light of the new solvency guidelines, is another path to be explored, as well as extending the model to the reinsurance industry.

Model or Scenario	Item	Expected Value	S.E.	$\mathrm{VaR}_{0.95}$	$\mathrm{VaR}_{0.995}$
Individual MSN	IBNR^{+}	$2,970,645$			
Theoretical	RBNS	$5,433,548$			
(until settlement)	Total	$8,404,192$			
Individual MSN	IBNR^{+}	$3,035,519$	$494,771$	$3,912,159$	$4,673,340$
Simulated	RBNS	$5,439,318$	$704,701$	$6,650,958$	$7,738,003$
(until settlement)	Total	$8,474,837$	$853,812$	$9,927,439$	$11,105,174$
Individual MSN Sim. + Unc. (until settlement)	Total	$8,533,066$	$875,989$	$10,054,807$	$11,219,311$
Individual MSN Sim. + Pol. Limit (until settlement)	Total	$8,464,661$	$823,752$	$9,875,418$	$10,912,072$
Individual MSN Sim. + Pol. Limit (until triangle bound)	Total	$7,131,164$	$766,852$	$8,449,221$	$9,353,716$
Mack Chain-Ladder	Total	$9,082,114$	$1,184,546$	$11,150,686$	$12,583,834$
Observed (bold, Table 2)	Total	$7,684,000$			

Table 5: Bodily Injury: comparison of estimation results. IBNR^{+} denotes the combination of IBNR and RBNP reserves. Results are displayed for: analytical best estimates (until settlement of each claim), corresponding simulation based results, simulation based results incorporating uncertainty in location parameters, simulation based results accounting for individual policy limit of 2.5 MEuro, simulation based results accounting for policy limits and developing until development year 8. Mack's chain–ladder results for Table 2 are displayed. Observed amount (i.e. sum of bold numbers in Table 2) is 7,684,000 euro.

Acknowledgements

The authors thank two anonymous referees and the associate editor for useful comments which helped to improve the paper substantially. Katrien Antonio acknowledges financial support from the Casualty Actuarial Society, Actuarial Foundation and the Committee on Knowledge Extension Research of the Society of Actuaries, and from NWO through a Veni 2009 grant.

Model or Scenario	Item	Expected Value	S.E.	$\text{VaR}_{0.95}$	$\text{VaR}_{0.995}$
Individual MSN	IBNR^+	$1,785,219$			
Theoretical	RBNS	$535,517$			
(until settlement)	Total	$2,320,735$			
Individual MSN	IBNR^+	$1,786,860$	$42,750$	$1,858,927$	$1,903,514$
Simulated	RBNS	$535,561$	$20,398$	$563,878$	$591,452$
(until settlement)	Total	$2,322,421$	$47,125$	$2,399,713$	$2,447,808$
Individual MSN Sim. + Unc. (until settlement)	Total	$2,345,550$	$49,842$	$2,424,624$	$2,473,893$
Individual MSN Sim. + Pol. Limit (until settlement)	Total	$2,318,058$	$46,489$	$2,404,582$	$2,447,490$
Individual MSN Sim. + Pol. Limit (until triangle bound)	Total	$2,312,532$	$46,786$	$2,388,427$	$2,431,461$
Mack Chain-Ladder	Total	$3,024,375$	$411,507$	$3,744,588$	$4,247,807$
Observed (bold, Table 3)	Total	$2,102,800$			

Table 6: Material Damage: comparison of estimation results. IBNR^+ denotes the combination of IBNR and RBNP reserves. Results are displayed for: analytical best estimates (until settlement of each claim), corresponding simulation based results, simulation based results incorporating uncertainty in location parameters, simulation based results accounting for individual policy limit of 2.5 MEuro, simulation based results accounting for policy limits and developing until development year 8. Mack's chain–ladder results for Table 3 are displayed. Observed amount (i.e. sum of bold numbers in Table 2) is 2,102,800 euro.

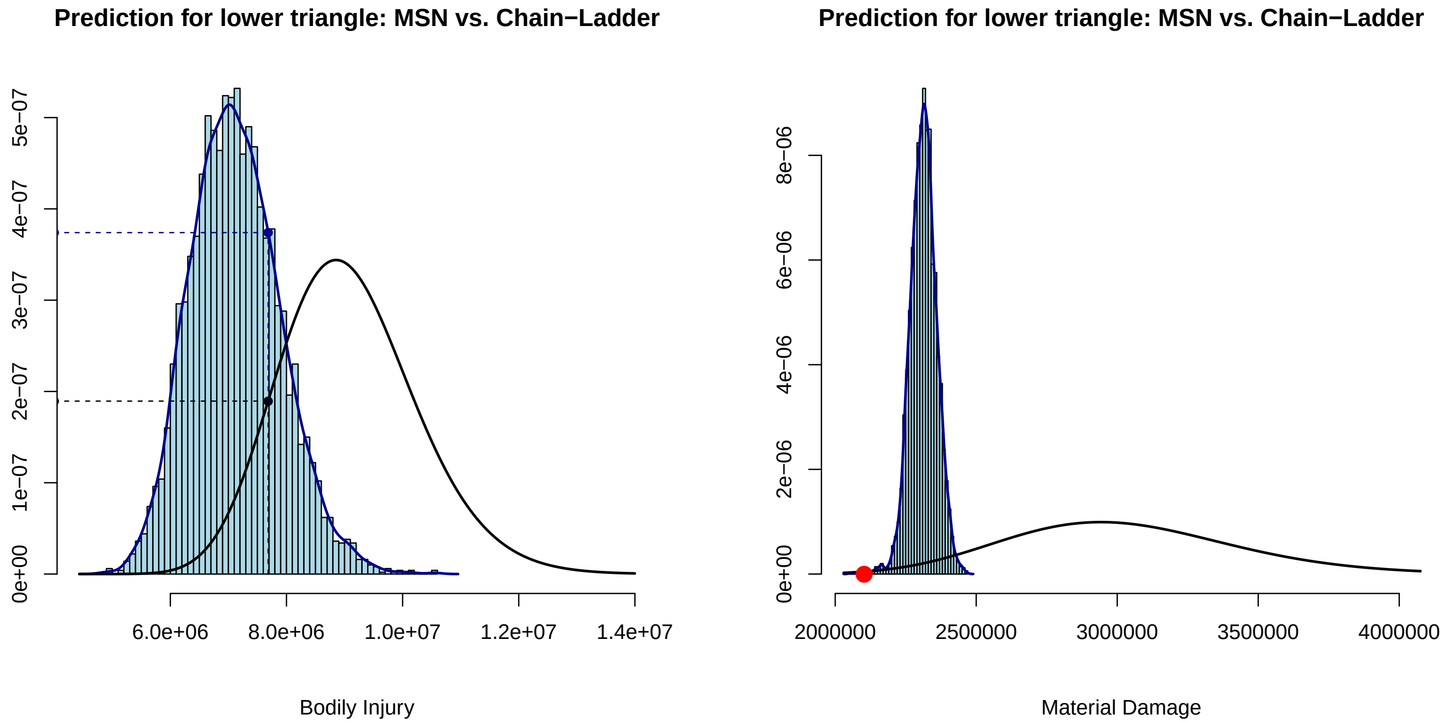


Figure 17: Histogram of the total reserve (light blue) obtained with the individual MSN model for Bodily Injury (left) and Material Damage (right). The histograms are based on 5,000 simulations (for BI) and 10,000 simulations (for MD) until the boundary of the triangle, taking the policy limit into account. The black reference line is the lognormal density function with mean and standard deviation as obtained with Mack's Chain–Ladder method. Dotted lines (on the BI plot) and red bullet (on the MD plot) represent the observed total payment for years 2005 to 2009 (August), i.e. the sum of the numbers in bold in Table 2 and 3.

References

K. Antonio and R. Plat. Micro–level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal, to appear*, 2012.

E. Arjas. The claims reserving problem in non–life insurance: some structural ideas. *ASTIN Bulletin*, 19(2):139–152, 1989.

A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178, 1985.

A. Deniz. *A class of multivariate skew distributions: properties and inferential issues*. PhD thesis, Bowling Green State University, Ohio, 2009.

P.D. England and R.J. Verrall. Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8:443–544, 2002.

A.K. Gupta and J.T. Chen. A class of multivariate skew normal models. *The Annals of the Institute of Statistical Mathematics*, 56(2):305–315, 2004.

S. Happ and M. Merz. Paid–incurred chain reserving method with dependence modelling. *ASTIN Bulletin, to appear*, 2012.

T. Mack. Distribution–free calculation of the standard error of chain–ladder reserve estimates. *ASTIN Bulletin*, 23(2):213–225, 1993.

T. Mack. The standard error of chain–ladder reserve estimates: recursive calculation and inclusion of a tail–factor. *ASTIN Bulletin*, 29:361–366, 1999.

M.D. Martinez, B. Nielsen, J.P. Nielsen, and R.J. Verrall. Cash flow simulation for a model of outstanding liabilities based on claim amounts and claim numbers. *ASTIN Bulletin*, 41(1):107–129, 2011.

M.D. Martinez, J.P. Nielsen, and R.J. Verrall. Double chain–ladder. *ASTIN Bulletin*, 42 (1):59–76, 2012.

M. Merz and M.V. Wüthrich. Pic claims reserving method. *Insurance: Mathematics and Economics*, 46(3):568–579, 2010.

R. Norberg. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bulletin*, 23(1):95–115, 1993.

R. Norberg. Prediction of outstanding liabilities II. Model extensions variations and extensions. *ASTIN Bulletin*, 29(1):5–25, 1999.

M. Pigeon and M. Denuit. Composite lognormal-pareto model with random threshold. *Scandinavian Actuarial Journal*, 3:177–192, 2011.

C. Roberts and S. Geisser. A necessary and sufficient condition for the square of a random variable to be gamma. *Biometrika*, 53(1/2):275–278, 1966.

R. Verrall, J.P. Nielsen, and A. Jessen. Including count data in claims reserving. *ASTIN Bulletin*, 40(2):871–887, 2010.

Mario V. Wüthrich and Michael Merz. *Stochastic claims reserving methods in insurance*. Wiley Finance, 2008.

X. B. Zhao, X. Zhou, and J. L. Wang. Semiparametric model for prediction of individual claim loss reserving. *Insurance: Mathematics and Economics*, 45(1):1–8, 2009.

A Proof of Proposition 2.2.

For a Multivariate Skew Symmetric random vector $((U+1)\times 1)$

$$\ln(\mathbf{\Lambda}_{U+1}) = \begin{bmatrix}\ln(Y_1) & \ln(\lambda_1) & \dots & \ln(\lambda_U)\end{bmatrix}' \tag{17}$$

and a $((U+1)\times 1)$ vector $\mathbf{t}$, the moment generating function is given by (see Deniz (2009))

$$\begin{aligned} & M_{\ln(\mathbf{\Lambda}_{U+1})}(\mathbf{t}) \\ = \; & \exp\left(\mathbf{t}_n'\boldsymbol{\mu}_{U+1}\right)\cdot E\left[\exp\left(\left(\mathbf{\Sigma}_{U+1}^{1/2}\right)'\mathbf{t}_n'\mathbf{z}_{U+1}\right)\cdot\prod_{j=1}^{U+1}H\left(\Delta_j\mathbf{e}_j'\mathbf{z}_{U+1}\right)\right]_{g^*(\mathbf{z}_{U+1})}. \end{aligned} \tag{18}$$

By definition,

$$\begin{aligned} M_{\ln(\mathbf{\Lambda}_{U+1})}(\mathbf{t}) &= E\left[\exp\left(\ln(\mathbf{\Lambda}_{U+1})'\,\mathbf{t}\right)\right] \\ &= E[\exp\left(\ln(Y_1)\,t_1 + \ln(\lambda_1)\,t_2 + \dots + \ln(\lambda_U)\,t_{U+1}\right)]. \end{aligned}$$

Taking $\mathbf{t} = \mathbf{t}_n = [\; n \quad n \quad \dots \quad n \;]'$ we obtain

$$\begin{aligned} M_{\ln(\mathbf{\Lambda}_{U+1})}(\mathbf{t}_n) &= E[\exp\left\{n\left(\ln(Y_1) + \ln(\lambda_1) + \dots + \ln(\lambda_U)\right)\right\}] \\ &= E\left[(Y_1\cdot\lambda_1\cdot\lambda_2\cdot\ldots\cdot\lambda_U)^n\right]. \end{aligned} \tag{19}$$

The n^{th} moment of an IBNR claim C is given by

$$\begin{aligned} E[C^n] &= E\left[E\left[(Y_1\cdot\lambda_1\cdot\lambda_2\cdot\ldots\cdot\lambda_u)^n\,|U=u\right]\right]_U \\ &= E\left[M_{\ln(\mathbf{\Lambda}_{U+1})}(\mathbf{t}_n)\right]_U \\ &= E\left[\exp\left(\mathbf{t}_n'\boldsymbol{\mu}_{U+1}\right)E\left[\exp\left(\left(\mathbf{\Sigma}_{U+1}^{1/2}\right)'\mathbf{t}_n'\mathbf{z}_{U+1}\right)\cdot\prod_{j=1}^{U+1}H\left(\Delta_j\mathbf{e}_j'\mathbf{z}_{U+1}\right)\right]_{g^*(\mathbf{z}_{U+1})}\right]_U. \end{aligned} \tag{20}$$

For the specific case of a Multivariate Skew Normal distribution, the result becomes

$$\begin{aligned} & E[C^n] \\ = \; & E\left[2^{U+1}\exp\left(\mathbf{t}_n'\boldsymbol{\mu}_{U+1} + 0.5\mathbf{t}_n'\boldsymbol{\Sigma}_{U+1}^{1/2}\left(\boldsymbol{\Sigma}_{U+1}^{1/2}\right)'\mathbf{t}_n\right)\cdot\prod_{j=1}^{U+1}\Phi\left(\frac{\Delta_j\left(\left(\boldsymbol{\Sigma}_{U+1}^{1/2}\right)'\mathbf{t}_n\right)_j}{\sqrt{1+\Delta_j^2}}\right)\right]_U. \end{aligned} \tag{21}$$

B Proof of Proposition 2.3.

The conditional Multivariate Skew Normal random vector defined by

$$\begin{aligned} & \ln\left(\boldsymbol{\Lambda}_B|\boldsymbol{\Lambda}_A = \boldsymbol{\ell}_A\right) \\ = \; & \left[\ln\left(y_1\right) \quad \ln\left(\ell_1\right) \quad \ldots \quad \ln\left(\ell_{u_A-1}\right) \quad \ln\left(\lambda_{u_A}\right) \quad \ldots \quad \ln\left(\lambda_U\right)\right] \end{aligned} \tag{22}$$

follows a Multivariate Skew Normal distribution with parameters $\boldsymbol{\mu}_{U+1}^*$, $\boldsymbol{\Sigma}_{U+1}^*$ and $\boldsymbol{\Delta}_{U+1}^*$ as defined in Proposition 2.3 (see Deniz (2009)). The rest of the proof is similar to the reasoning given in Section A.

C Proof of Proposition 2.4

(a) For IBNR claims, the expected value of the total claim amount is

$$E[\text{IBNR}|\mathcal{I}] = E\left[\sum_{i=1}^{I}\sum_{k=1}^{K_{\text{IBNR},i}} Y_1^{(ik)}\cdot\lambda_1^{(ik)}\cdot\ldots\cdot\lambda_{U_{ik}}^{(ik)}\right], \tag{23}$$

where $K_{\text{IBNR},i}$ is the random variable representing the number of IBNR claims from occurrence period i. Because $K_{\text{IBNR},i}$ and $\boldsymbol{\Lambda}_{U+1}$ are independent, we obtain

$$\begin{aligned} E[\text{IBNR}|\mathcal{I}] &= \sum_{i=1}^{I} E[K_{\text{IBNR},i}]\, E\left[Y_1^{(ik)}\cdot\lambda_1^{(ik)}\cdot\ldots\cdot\lambda_{U_{ik}}^{(ik)}\right] \\ &= E[K_{\text{IBNR}}]\cdot E\left[Y_1^{(ik)}\cdot\lambda_1^{(ik)}\cdot\ldots\cdot\lambda_{U_{ik}}^{(ik)}\right]. \end{aligned} \tag{24}$$

The result than follows from Proposition 2.2. The proof is similar for RBNP claims.

(b) For RBNS claims, the expected value of the total claim amount is

$$
\begin{aligned}
E[\mathrm{RBNS}|\mathcal{I}] &= \sum_{(ik)_{\mathrm{RBNS}}} E\left[y_1^{(ik)} \cdot \ell_1^{(ik)} \cdot \ldots \cdot \ell_{u_A^{(ik)}-1}^{(ik)} \cdot \lambda_{u_A^{(ik)}}^{(ik)} \cdot \ldots \cdot \lambda_{U_{ik}}^{(ik)} \right] \\
&= \sum_{(ik)_{\mathrm{RBNS}}} y_1^{(ik)} \cdot \ell_1^{(ik)} \cdot \ldots \cdot \ell_{u_A^{(ik)}-1}^{(ik)} \cdot E\left[\lambda_{u_A}^{(ik)} \cdot \ldots \cdot \lambda_{U^{(ik)}}^{(ik)} \right]. \qquad (25)
\end{aligned}
$$

The proof then follows from Proposition 2.3.