

Generalized Mack Chain-Ladder Model of Reserving with Robust Estimation

by Przemyslaw Sloma

ABSTRACT

In this paper we consider the problem of stochastic claims reserving in the framework of development factor models (DFM). More precisely, we provide the generalized Mack chain-ladder (GMCL) model that expands the approaches of Mack (1993; 1994; 1999), Saito (2009) and Murphy, Bardis, and Majidi (2012). Our general flexible tool of reserving provides the solution to the one of the major challenges of day-to-day actuarial practice, which is quantifying the variability of reserves in the case where different methods of selecting loss developments factors (LDFs) are applied. We develop the theoretical background to estimate the conditional mean square error of prediction (MSEP) of claims reserves that is consistent with actuarial practice in selecting the LDFs.

Moreover, we present an example of GMCL's application in which we indicate how to bridge the estimation of parameters in the chain-ladder framework with the robust estimation techniques. Finally, we show how our approach can be used in validation of the reserve risk evaluation in the Solvency 2 context.

KEYWORDS

*Non-life insurance, stochastic claims reserving,
Mack chain ladder, robust estimation solvency 2*

1. Introduction and motivation

The provision for outstanding claims is one of the main components of technical provisions of insurance company's liabilities. Measuring the deviation of the true amount of reserves from its estimation is one of the major actuarial challenge. Senior managers, shareholders, rating agencies, and insurance regulators all have an interest in knowing the magnitude of these potential variations (reserve uncertainty) since companies with large potential deviations need more capital or reinsurance.

One of the most known method of reserving used in practice is the approach called chain-ladder. This method belongs to the family of development factor models (DFM). The first stochastic approach based on the chain-ladder technique was proposed by Mack (1993; 1994). In these studies, Mack proposed the estimation of the mean square error of prediction (MSEP) of claims reserves based upon all-year volume-weighted average of loss development factors (also called: link ratios, age-to-age factors, report-to-report factors). The variance structure was supposed to be proportional to the development period's initial loss. This assumption is sufficient for the weighted average development factors to have optimal statistical properties (BLUE-best linear unbiased estimate). Some authors (see Murphy 1996, p. 188; Mack 1999, p. 15) pointed out that the estimation of chain-ladder factors is connected with the estimation in the framework of linear model by weighted least squares (WLS) regression approach. They also observed that by modifying the original variance assumption from Mack (1993) the corresponding estimators of chain-ladder development factors keep their BLUE property (see Murphy 1996; Barnett and Zehnwirth 2000; Saito 2009). It is worth underlying that the modification of variance assumption leads to different point estimators for development factors (arithmetic average, slope of regression etc.; see Remark 3.1 for more details).

One of the major challenges in everyday actuarial practice is selecting the loss development factors.

The adjustments to make data more homogeneous are often justified for number of reasons: unstable run-off triangles, outliers, inaccurate and incomplete data, etc.). Most actuaries use somewhat arbitrary rules of thumb in selecting the loss development factors (LDFs). Blumsohn and Laufer (2009) describe this topic in great detail. In this project, a group of actuaries were asked to select LDFs for an incurred run-off triangle. The important number of ways of LDFs selection was provided by the participants. The approaches proposed to evaluate the estimation of expected value of reserves varied widely and the additional information about the error of prediction of this estimation could be helpful in decision making. That is why it is extremely important from a practical point of view to have a method that provides the estimation of conditional MSEP of ultimate claims (or reserves) in the context of LDFs selection by actuaries. In the present study we provide such a tool embedded in the theoretical framework to quantify the standard error of prediction of the claims reserves in the case where some factors have been excluded from the estimation of model parameters. However, we do not judge whether these ad hoc approaches of selecting factors are correct or wrong. We rather assume that the expert judgment taken by an actuary could always be justified by his specific knowledge of considered business.

Measuring the variability of the reserves in this context is poorly developed in the literature. That is why in practice actuaries and reserving software developers often use the proxy methods based on formula for MSEP derived in Mack (1993). The approximations mainly consist of replacing the main parameters by their estimators computed by the other approach without changing the main formula. This procedure is incorrect because in the chain-ladder framework the formula for MSEP depends among others on the standard error of chain-ladder factors and it is not accurate to simply plug-in the new estimators in the old formula. The other proxy method often used in practice consists on applying the coefficients of variation of ultimate loss from Mack (1993)

(ratio of square root of MSEP of ultimate loss over ultimate loss) in order to derive the MSEP estimators of a new approach. It turns out that in general these approximations are highly inappropriate (see example in Section 6.3).

We think that, in some simple cases (no curve fitting for LDFs, for example) the approximations mentioned above are the consequences of bad understanding of the main formula for estimation of MSEP in Mack (1993). Moreover, the approximations used by actuaries and actuarial software developers could be avoided by using the more appropriate existing models. One such model was proposed by Mack (1999). To our knowledge, this was the first study that showed how to measure the uncertainty of reserves in the situation when an actuary selects the LDFs. In our opinion, the important results obtained in this paper are not always used in practice because, instead of explicit formula, the recursive equation is given there for estimation of MSEP of ultimate loss.

Mack (1999) is an important paper that allows the fully understanding of the MSEP formula and avoid the inappropriate approximation when it is not necessary. We summarize the details of this method in Section 3. One of the major limitations of this method is the underestimation the MSEP of ultimate loss in the case where the number of excluding data is important. We discuss this topic in detail in Section 3.5. One possible solution to overcome this difficulty is to extend the existing approach proposed by Mack (1999).

Therefore, we propose a general approach for stochastic claims reserving in the framework of chain-ladder model, extending the model proposed by Mack (1999) and Murphy, Bardis, and Majidi (2012). This extension is three fold. First, our general tool has a educational role and makes it possible to validate the results from other approaches. More precisely, our general formula for estimation of MSEP of outstanding loss liabilities can be used to fully understand the Mack (1993), Mack (1994), and Mack (1999) model. Furthermore, under new solvency requirements of Solvency II, insurance companies use the bootstrap-

type stochastic reserving methods to determine the economic capital corresponding to the *reserve risk*. The bootstrap method allows estimation of a whole claims reserves distribution via resampling techniques and Monte Carlo simulations. It seems to be crucial for non-life insurance companies to be able to validate the results given by the industrial software where we do not have access to the code and when the number of shortcuts may be applied. Our approach can be used to validate the estimation of the first two moments of the loss distribution in the case where selection of development factors was employed and the different weights in estimation of chain-ladder factors and volatility parameters were used (see Section 6.3 for more details).

Second, our general Mack chain-ladder (GMCL) model can be used to construct the proxy solutions to overcome the limits of Mack's (1999) approach, i.e., the use of the same weights for parameters estimation (see discussion in Section 3.6). This means that, for the methods where we eliminate the considerable number of observations, we reduce as well the data for variability of the reserves. This mechanically impacts the estimation of MSEP of loss liabilities, which in such cases is generally underestimated. We propose then the possible solution to overcome this kind of difficulty (see Section 6.1 for more details).

Finally, the third and really important application from a practical point of view consists of bridging the point estimation of chain-ladder parameters with the theory of robust statistics. As mentioned above, the point estimators of chain-ladder factors can be obtained in the linear regression framework by applying the weighted least squares procedure. It is well known that the OLS estimators are fragile to the outliers. That is why we propose using the robust techniques of estimation such as: M-estimators, L^p -estimators, etc (see Section 6.2).

The reminder of this paper is organized as follows. In Section 2 we present our notations and definitions. We review in Section 3 the MCL and its main limitations. In Section 4, we present the (GMCL) and the main results are derived in Section 5. Finally,

Section 6 introduces the numerical applications of GMCL. All proofs are provided in the Appendix. The related topics such as: tail factor, curve fitting, diagnostics and validation of the main model hypothesis, are out of scope of this paper and will be treated elsewhere.

2. Notations and definitions

2.1. Run-off triangle

Let C_{ij} denote the random variables (cumulative payments, incurred, reported claims numbers, etc.) for accident year $i \in \{1, \dots, I\}$ until development year $j \in \{1, \dots, J\}$, where the accident year is referred to as the year in which an event triggering insurance claims occurs. We assume that C_{ij} are random variables observable for calendar years $i + j \leq I + 1$ and non-observable (to be predicted) for calendar years $i + j > I + 1$. The observable C_{ij} are represented by the so-called run-off trapezoids ($I > J$) or run-off triangles ($I = J$). Table 1 gives an example of a typical run-off triangle. In order to simplify our notation, we assume that $I = J$ (run-off triangle). However, all the results we present here can be easily extended to the case when the last accident year for which data is available is greater than the last development year, i.e., $I > J$ (run-off trapezoid).

2.2. Outstanding reserves

Let R_i et R denote the *outstanding claims liabilities* for accident year $i \in \{1, \dots, I\}$,

Table 1. Run-off triangle ($I = J$)

Accident Year i	Development Year j					
	1	2	3	4	j	$\dots$ J
1						
2						
3						
$I - j$						
$I - 2$						
$I - 1$						
I						

C_{ij}
(observations)

C_{ij}
(to be predicted)

$$R_i = C_{i,I} - C_{i,I-i+1}, \quad (2.1)$$

and the total outstanding loss liabilities for all accident years,

$$R = \sum_{i=1}^I R_i. \quad (2.2)$$

We use the term *claims reserves* to describe the prediction of the outstanding loss liabilities. Hence, let $\hat{R}_i$ and R denote the claims reserves for accident year i , $\hat{R}_i = \hat{C}_{i,I} - C_{i,I-i+1}$, $i \in \{1, \dots, I\}$, and the total claims reserves for aggregated accident years, $\hat{R} = \sum_{i=1}^I \hat{R}_i$, respectively, where $\hat{C}_{i,I}$ is a predictor for $C_{i,I}$.

2.3. (Conditional) mean square error of prediction (MSEP)

As already stated above, finding suitable prediction of ultimate loss is rather the beginning of the process of reserving, and insurers need to assess the variability of these amounts. We are interested then in the quantification of the prediction uncertainty of the ultimate loss, i.e., $\hat{C}_{i,I}$ and $\sum_{i=1}^I \hat{C}_{i,I}$ (or equivalently of *claims reserves*, i.e., $\hat{R}_i$ and $\hat{R} = \sum_{i=1}^I \hat{R}_i$). For that, we have to choose an appropriate risk measure which determines a conception of measuring the “distance” between the prediction and the actual outcomes. In this paper, following the actuarial literature, we quantify the prediction uncertainty using the most popular such measure, the so-called mean-square error of prediction (MSEP).

$$mse_{\hat{C}_{i,I}|D_I}(C_{i,I}) = E \left[\left(\hat{C}_{i,I} - C_{i,I} \right)^2 \middle| D_I \right], \quad (2.3)$$

$$mse_{\sum_{i=1}^I \hat{C}_{i,I}|D_I} \left(\sum_{i=1}^I C_{i,I} \right) = E \left[\left(\sum_{i=1}^I \hat{C}_{i,I} - \sum_{i=1}^I C_{i,I} \right)^2 \middle| D_I \right], \quad (2.4)$$

where

$$D_I = \{C_{i,j} : i + j \leq I + 1\}, \quad (2.5)$$

denote the claims data available at time $t = I$.

3. Mack chain-ladder (MCL) model

A major everyday challenge of actuarial work is selecting loss development factors for number of reasons (outliers in triangle, inaccurate data, incompleteness, etc). Most actuaries use somewhat arbitrary rules of thumb in selecting the loss ratios. In Blumsohn and Laufer (2009), a group of actuaries were asked to select age-to age factors for a 12-years triangle of umbrella business. The important number of ways of selecting loss ratios was provided by the participants. It is important, then, from practical point of view to have a method that provides the estimation of conditional MSEP of ultimate claims in the context of factor selection of actuaries.

To the best of our knowledge, the paper by Mack (1999) is one of the first studies dealing with factors selection and variability of reserves estimation in the framework of the chain-ladder method. This paper is an extension of Mack (1993).

In the remaining part of this section we recall the assumptions of the Mack chain-ladder (MCL) model from Mack (1999). Afterwards, we present the numerical example illustrating the limits of this approach. Finally, we indicate the possible expansion of MCL method and its potential applications.

3.1. Model assumptions of MCL method

Let define the individual development factors, for $1 \leq i \leq I-1$ and $1 \leq k \leq I-1$,

$$F_{i,k} = C_{i,k+1}/C_{i,k}. \quad (3.1)$$

Following Mack (1999), we assume [(MCL.1)]

1. There exist constants $f_k > 0$ such that

$$E(F_{i,k} | C_{i,1}, \dots, C_{i,k}) = f_k.$$

The parameters f_k are often called *loss development factors* (LDF), *link ratios* or *age-to-age factors*.

2. There exist constants $\sigma_k^2 > 0$ such that for all $1 \leq i \leq I$ and $1 \leq k \leq I-1$ we have

$$\begin{aligned} \text{Var}(F_{i,k} | C_{i,1}, \dots, C_{i,k}) &= \frac{\sigma_k^2}{w_{i,k} C_{i,k}^\alpha}, \\ \text{with } w_{i,k} &\in [0, 1]. \end{aligned} \quad (3.2)$$

The parameters σ_k are referred here as *variance parameters* (LDF).

3. The accident years $(C_{i,1}, \dots, C_{i,I})_{1 \leq i \leq I}$ are independent.

3.2. Estimation of parameters in the MCL model

- Given the information D_I and for $1 \leq k \leq I-1$, the factors f_k are estimated by

$$\hat{f}_k = \frac{\sum_{i=1}^{I-k} w_{i,k} C_{i,k}^\alpha F_{i,k}}{\sum_{i=1}^{I-k} \gamma_{i,k}}, \quad \alpha \in \{0, 1, 2\}. \quad (3.3)$$

- Given the information D_I and for $1 \leq k \leq I-2$, the variance parameters σ_k^2 are estimated by

$$\hat{\sigma}_k^2 = \frac{1}{I_k - 1} \sum_{i=1}^{I-k} w_{i,k} C_{i,k}^\alpha (F_{i,k} - \hat{f}_k)^2, \quad \alpha \in \{0, 1, 2\}, \quad (3.4)$$

where I_k represents the number of weights $w_{i,k}$ different from 0, namely, $I_k := \text{card}\{i: w_{i,k} \neq 0\}$.

Formula (3.4) does not yield an estimator for $\hat{\sigma}_{I-1}^2$ because it is not possible to estimate this parameter from the single observation $C_{I,I}/C_{I,I-1}$. Following Mack (1993; 1994; 1999), if $f_{I-1} = 1$ and if the claims development is believed to be finished after $I-1$ years we can put $\hat{\sigma}_{I-1}^2 = 0$. If not, the simple formula of extrapolation can be applied by requiring $\hat{\sigma}_{I-3}/\hat{\sigma}_{I-2} = \hat{\sigma}_{I-2}/\hat{\sigma}_{I-1}$. This leads to the following definition

$$\hat{\sigma}_{I-1}^2 := \min(\hat{\sigma}_{I-2}^4/\hat{\sigma}_{I-3}^2, \min(\hat{\sigma}_{I-3}^2, \hat{\sigma}_{I-2}^2)). \quad (3.5)$$

Remark 3.1. The parameter α determines the different ways of estimation of f_k . For the sake of simplicity, let us assume that $w_{ij} = 1$ for all i, j .

We present below the possible choices of α and their interpretation.

1. If $\alpha = 1$ we get the classical chain ladder estimate of f_k

$$\hat{f}_k = \frac{\sum_{i=1}^{I-k} C_{i,k} F_{i,k}}{\sum_{i=1}^{I-k} C_{i,k}} = \frac{\sum_{i=1}^{I-k} C_{i,k+1}}{\sum_{i=1}^{I-k} C_{i,k}}, \quad \text{for } 1 \leq k \leq I-1.$$

2. If $\alpha = 0$ we get the model for which the estimators of the age-to-age factors f_k are the straightforward average of the observed individual development factors $F_{i,j}$ defined via (3.1), i.e.,

$$\hat{f}_k = \frac{1}{I-k} \sum_{i=1}^{I-k} F_{i,k}, \quad \text{for } 1 \leq k \leq I-1.$$

3. If $\alpha = 2$ we get the model for which the estimators of the age-to-age factors f_k are the results of an ordinary regression of $\{C_{i,k+1}\}_{i \in \{1, \dots, I-k-1\}}$ against $\{C_{i,k}\}_{i \in \{1, \dots, I-k\}}$ with intercept 0, i.e.,

$$\hat{f}_k = \frac{\sum_{i=1}^{I-k} C_{i,k}^2 F_{i,k}}{\sum_{i=1}^{I-k} C_{i,k}^2} = \frac{\sum_{i=1}^{I-k} C_{i,k} C_{i,k+1}}{\sum_{i=0}^{I-k-1} C_{i,k}^2}, \quad \text{for } 0 \leq k \leq I-1.$$

3.3. Properties of estimators from MCL model

Proposition 3.1

- i. The estimators $\hat{f}_k$ given in (3.3) are unbiased and uncorrelated.
- ii. The estimators $\hat{f}_k$ of f_k have the minimal variance among all unbiased estimators of f_k which are the weighted average of the observed development factors $F_{i,k}$.
- iii. The estimator $\hat{\sigma}_k^2$, given in (3.4) is the unbiased estimator of the parameter σ_k^2 .
- iv. Under the model assumptions (MCL.1) and (MCL.3) we have

$$E(C_{i,I} | D_I) = C_{i,I+1-i} f_{i,I+1-i} \cdot \dots \cdot f_{I-1}.$$

This implies, together with the fact that $\hat{f}_k$ are uncorrelated, that $\hat{C}_{i,I}$ is unbiased estimator of $E(C_{i,I} | D_I)$.

- v. The expected values of the estimator

$$\hat{C}_{i,I} = C_{i,I+1-i} \cdot \prod_{k=I+1-i}^{I-1} \hat{f}_k,$$

for the ultimate claims amount and of the true ultimate claims amount $C_{i,I}$ are equal, i.e., $E(\hat{C}_{i,I}) = E(C_{i,I})$, $2 \leq i \leq I$.

The proof is postponed to Appendix A.3.

3.4. Estimators of conditional MSEP in MCL model

3.4.1. Single accident years

Under assumptions of the MCL model we have the following estimator for the conditional estimation error of a single accident year $i \in \{2, \dots, I\}$:

$$\widehat{mse}_{\hat{C}_{i,I} | D_I}(C_{i,I}) = (\hat{C}_{i,I})^2 \cdot \sum_{k=I-i+1}^{I-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \left(\frac{1}{\hat{w}_{i,k} \hat{C}_{i,k}^\alpha} + \frac{1}{\sum_{j=1}^{I-k} w_{j,k} C_{j,k}^\alpha} \right), \quad (3.6)$$

where, for $i+k > I+1$, we define $\hat{w}_{i,k} := 1$ and $\hat{f}_j$ and $\hat{\sigma}_j^2$ are given in (3.3) and (3.4)–(3.5) respectively.

3.4.2. Aggregated accident years

$$\widehat{mse}_{\sum_{i=1}^I \hat{C}_{i,I} | D_I} \left(\sum_{i=1}^I C_{i,I} \right) = \sum_{i=2}^I \widehat{mse}_{\hat{C}_{i,I} | D_I}(C_{i,I}) + \sum_{i=2}^I \hat{C}_{i,I} \left(\sum_{j=i+1}^I \hat{C}_{j,I} \right) \sum_{k=I-i+1}^{I-1} 2 \frac{\hat{\sigma}_k^2 / (\hat{f}_k)^2}{\sum_{l=1}^{I-k} w_{l,k} C_{l,k}^\alpha}, \quad (3.7)$$

where $\hat{f}_j$ and $\hat{\sigma}_j^2$ are given in (3.3) and (3.4)–(3.5) respectively.

3.5. Numerical application of the MCL method

As mentioned above, the factors selection methods are an integral part of everyday actuarial practice.

Here we choose from Blumsohn and Laufer (2009) several such methods where estimates are computed as different averages using varying weights and varying number of accident years: all/3/5-years weighted average and all excluding higher and lower (AEHL) factor average. We consider as well other popular methods in actuarial practice based on sample median.

More precisely, for RAA run-off triangle (see Appendix B, Section B.8), we apply the MCL model from Mack (1999) with the following parameters. For all five methods described below we choose $\alpha = 0$ ($\hat{f}_k$ arithmetic averages of $F_{i,k}$) and we compute the estimators $\hat{f}_k$ and $\hat{\sigma}_k^2$ according to formula (3.3) and (3.4)–(3.5), respectively. This allows us to compare the results with *sample Median* method which is rather consistent with straightforward average of development factors (see method number (5) below)

- (1) **ALL AV:** $\hat{f}_k$ are computed as arithmetic average of all individual link ratios $F_{i,j}$. More precisely, we define the weights in the following way: $w_{i,j} = 1$ for all i, j .
- (2) **AEHL:** $\hat{f}_k$ are computed as arithmetic average of all individual link ratios, excluding the highest and the lowest values of $F_{i,j}$. More precisely, we define the weights in the following way: for fixed j , $w_{i,j} = 0$ for i such that $F_{i,j} = F_{(I-j),j}$ and $F_{i,j} = F_{(1),j}$, where $F_{(k),j}$ for $k = 1, \dots, I-j$ denotes the order statistics of $F_{i,j}$. For remaining indices i , for fixed j , we take $w_{i,j} = 1$.
- (3) **5 Years AV:** $\hat{f}_k$ are computed as an arithmetic average of individual link ratios $F_{i,j}$ from five latest accidents years. More precisely, we define the weights in the following way: $w_{i,j} = 1$ for $i = I-j, \dots, I-j-4$. For remaining indices i , for fixed j , we take $w_{i,j} = 0$.
- (4) **3 Years AV:** $\hat{f}_k$ are computed as an arithmetic average of individual link ratios $F_{i,j}$ from three latest accidents years. More precisely, we define the weights in the following way: $w_{i,j} = 1$ for $i = I-j, \dots, I-j-2$. For remaining indices i , for fixed j , we take $w_{i,j} = 0$.

- (5) **Median:** $\hat{f}_k$ are computed as an arithmetic average of individual link ratios $F_{i,j}$ in the way to obtain the *sample median*. More precisely, we put $w_{i,j} = 1$ or $w_{i,j} = 0$ in the way that the estimators of the age-to-age factors f_k are given by

$$\hat{f}_k = \text{median}\{F_{i,k} : i \in \{1, \dots, I-k\}\}.$$

The *median* denotes the sample median that for the sample $X_1, \dots, X_n$ is computed by

$$\text{median}\{X_i : i \in \{1, \dots, n\}\} := \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd} \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} & \text{otherwise} \end{cases}.$$

where $X_{(k)}$ denotes the k -th order statistics of the sample $X_1, \dots, X_n$.

Remark 3.2. In the case where there is only one observation in estimation of parameter σ_k (odd number of data in sample median computation) we choose the additional $F_{i,k}$ factor in order to have two observations and be able to apply the formula (3.4).

In Table 2, we present the estimation of total amount of claims reserves $\hat{R}$ as well as the value of estimators of aggregated $MSEP(\hat{R})$. Recall that $\hat{R} := \sum_{i=1}^I \hat{R}_i$, where $\hat{R}_i := \hat{C}_{i,I} - C_{i,I+1}$. We observe that to obtain $\hat{R}$ it is enough to have the estimators $C_{i,I}$ of ultimate claims $\hat{C}_{i,I}$ for all accident year i . In consequence $MSEP(\hat{R}) = MSEP(\sum_{i=1}^I \hat{C}_{i,I})$ and we use the formula (3.7) to estimate this quantity. We compute as well the coefficient of variation of $\hat{R}$, given by $CV(\hat{R}) = \hat{R}/MSEP(\hat{R})^{1/2}$. The last two lines of Table 2 indicate the relative proportion of $\hat{R}$ and $MSEP(\hat{R})^{1/2}$, for each of five methods considered, in comparison to the *ALL AV* method which is the reference method in our example.

3.6. Limits of MCL method

As can be seen in Table 2, the four last methods (columns (2)–(5)) reduce significantly the estimation

Table 2. Estimation of total amount of outstanding loss liabilities ($\hat{R}$), value of estimator of aggregated $MSEP(\hat{R})^{1/2}$ and coefficient of variation $CV(\hat{R})$, for five methods

alpha=0					
Item/method	ALL AV (1)	AEHL (2)	5 Years AV (3)	3 Years AV (4)	Median (5)
$\hat{R}$	93 643	65 868	75 886	68 645	54 059
$MSEP(\hat{R})^{1/2}$	92 549	21 015	27 486	29 493	14 786
$CV(\hat{R})$	99%	32%	36%	43%	27%
	(1)/(1)	(2)/(1)	(3)/(1)	(4)/(1)	(5)/(1)
$\hat{R}$ (%)	100%	70%	81%	73%	58%
$MSEP(\hat{R})^{1/2}$ (%)	100%	23%	30%	32%	16%

of $MSEP(\hat{R})^{1/2}$ comparing to the first method *ALL AV*. For the methods (3), (4) and (5), this is mainly due to the elimination of relatively significant number of development factors from estimation especially for the first development years which correspond to the columns of the run-off triangle. This phenomena is especially seen in the case of *sample median* method in which, for each development factor we keep at most two of link ratios $F_{i,j}$ in estimation of f_k . From statistical point of view, this is clearly not enough to perform the robust estimation. As a consequence, this kind of methods reduce unnaturally the variability of reserves. This could be dangerous for example in terms of evaluation of the economical capital for reserve risk required by the new Solvency II regime.

Beyond the limits stated above, there are some incoherences with application of weights $w_{i,k}$ for the *AEHL* and *sample median* methods. Indeed, the weights $w_{i,k}$ should be $C_{i,k}$ measurable random variables in order to be able to derive the main results of MCL approach (see, for example, Proposition A.2). Although for the method *5 Year AV* and *3 Years AV* we can fix the weights without knowing the information D_i (knowing all observation in the run-off triangle, see (5)), this is not a case for the *AEHL* and *sample median* methods. The reason is that we need to know the observation $F_{i,k}$ in order to specify the corresponding weights for those two methods. That is why the weights $w_{i,k}$ are not $C_{i,k}$ measurable but rather D_i -measurable. This means that the formula

for the expectation of $\hat{f}_k$ and $Var(\hat{f}_k|B_k)$ are not correct. Regarding the *sample median* method, the derivation of $Var(\hat{f}_k|B_k)$ requires the computation of the moments of order statistics (see Jeng 2010) and those are strongly related to the distribution of $F_{i,j}$. To overcome these difficulties we propose two solutions: the simple *Proxy method* (see Section 6.1) and the more complex one based on a robust estimation (see Section 6.2). The first approach is programmed to avoid artificial volatility increase and it is based on all link ratios in estimation of volatility parameters σ_k (scale parameters in linear regression). The second method consists on developing an approach that allows using any robust estimators of f_k (location) and σ_k (scale) parameters.

4. General Mack chain-ladder model

4.1. Model assumptions

Before stating the main assumptions of our general approach, let us assume that functions $g_{\delta,j}: [0, \infty) \rightarrow [0, \infty)$ are Borel measurable. Let $\delta_{i,j}$ be the *non-negative* random variables defined by, $\delta_{i,j} := g_{\delta,j}(C_{i,j})$.

Our model is formalized by the following assumptions:

(GMCL.1) There exist constants $f_k > 0$ such that

$$E(F_{i,k} | C_{i,1}, \dots, C_{i,k}) = f_k.$$

(GMCL.2) There exist constants $\sigma_k^2 > 0$ such that for all $1 \leq i \leq I$ and $1 \leq k \leq I-1$ we have

$$\text{Var}(F_{i,k} | C_{i,1}, \dots, C_{i,k}) = \begin{cases} \frac{\sigma_k^2}{\delta_{i,k}} & \text{if } \delta_{i,k} \neq 0 \text{ a.s.,} \\ \infty & \text{if } \delta_{i,k} = 0 \text{ a.s.,} \end{cases} \quad (4.1)$$

where a.s. means *almost surely*.

(GMCL.3) The accident years $(C_{i,1}, \dots, C_{i,I})_{1 \leq i \leq I}$ are independent

We observe that from the above assumptions the main difference between MCL and GMCL lies in the variance assumption. This modification allows us to introduce different weights in estimation of the parameters f_k and σ_k .

4.2. Model estimators

Suppose that functions $g_{\gamma_j}: [0, \infty) \rightarrow [0, \infty)$ are Borel measurable. Let γ_{ij} be the *non-negative* random variables defined by, $\gamma_{ij} := g_{\gamma_j}(C_{i,j})$.

- Given the information D_i , the factors f_k are estimated by

$$\hat{f}_k = \frac{\sum_{i=1}^{I-k} \gamma_{i,k} F_{i,k}}{\sum_{i=1}^{I-k} \gamma_{i,k}}, \quad \text{for } 1 \leq k \leq I-1 \quad (4.2)$$

It becomes obvious from assumption (GMCL.2) that in order to compute correctly the variance of $\hat{f}_k$ (see Proposition A.2 in Appendix) we have to assume that

$$\{\text{if } \delta_{i,j} = 0 \text{ then } \gamma_{i,j} = 0\}. \quad (4.3)$$

- Given the information D_I , the variance parameters σ_k^2 are estimated by

$$\hat{\sigma}_k^2 = \frac{1}{I_k - 1} \sum_{i=1}^{I-k} \delta_{i,k} (F_{i,k} - \hat{f}_k)^2, \quad \text{for } 1 \leq k \leq I-2, \quad (4.4)$$

where I_k represents the number of weights $\delta_{i,k}$ different from 0, namely, $I_k := \text{card}\{i: \delta_{i,k} \neq 0\}$.

In the analogue way to (3.4) we define

$$\hat{\sigma}_{I-1}^2 = \min(\hat{\sigma}_{I-2}^4 / \hat{\sigma}_{I-3}^2, \min(\hat{\sigma}_{I-3}^2, \hat{\sigma}_{I-2}^2)). \quad (4.5)$$

Proposition 4.1.

- The estimators $\hat{f}_k$ given in (4.2) are unbiased and uncorrelated.
- For $k = 1, \dots, I-1$, if $\delta_{i,k} = \gamma_{i,k}$ for all i , then the estimators $\hat{f}_k$ of f_k have the minimal variance among all unbiased estimators of f_k which are the weighted average of the observed development factors $F_{i,k}$.

For $k = 1, \dots, I-1$, if $\delta_{i,k} \neq \gamma_{i,k}$ for some i , then the relative efficiency of $s.e.(\hat{f}_k^{\gamma \neq \delta} | B_k)$ with respect to $s.e.(\hat{f}_k^{\gamma = \delta} | B_k)$, i.e., the ratio

$$\frac{s.e.(\hat{f}_k^{\gamma \neq \delta} | B_k)}{s.e.(\hat{f}_k^{\gamma = \delta} | B_k)} = \frac{\text{Var}(\hat{f}_k^{\gamma \neq \delta} | B_k)^{1/2}}{\text{Var}(\hat{f}_k^{\gamma = \delta} | B_k)^{1/2}} = \frac{\sum_{j=1}^{I-k} \frac{\gamma_{j,k}^2}{\delta_{j,k}} \cdot \mathbf{1}_{\{\delta_{j,k} \neq 0\}}}{\sum_{j=1}^{I-k} \gamma_{j,k} \cdot \mathbf{1}_{\{\delta_{j,k} \neq 0\}}}$$

- For $k = 1, \dots, I-1$, if $\delta_{i,j} = \gamma_{i,j}$ for all i , then the estimator $\hat{\sigma}_k^2$, given in (4.4) is the unbiased estimator of the parameter σ_k^2 .

For $k = 1, \dots, I-1$, if $\delta_{i,k} \neq \gamma_{i,k}$ for some i , then the bias of the estimator $\hat{\sigma}_k^2$ is given by the following formula

$$E[\hat{\sigma}_k^2 - \sigma_k^2] = \frac{\sigma_k^2}{I_k - 1} E \left[\frac{\sum_{i=1}^{I-k} \delta_{i,k} \left(\frac{\sum_{j=1}^{I-k} \gamma_{j,k}^2 \cdot \mathbf{1}_{\{\delta_{j,k} \neq 0\}}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k} \right)^2} \right) - 1}{\left(\sum_{j=1}^{I-k} \gamma_{j,k} \right)^2} \right].$$

- Under the model assumptions (GMCL.1) and (GMCL.3) we have

$$E(C_{i,I} | D_I) = C_{i,I+1-i} f_{i,I+1-i} \cdot \dots \cdot f_{I-1}.$$

This implies, together with the fact that $\hat{f}_k$ are uncorrelated, that $\hat{C}_{i,I}$ is unbiased estimator of $E(C_{i,I} | D_I)$.

(v) The expected values of the estimator

$$\hat{C}_{i,I} = C_{i,I+1-i} \cdot \prod_{k=I+1-i}^{I-1} \hat{f}_k,$$

for the ultimate claims amount and of the true ultimate claims amount $C_{i,I}$ are equal, i.e., $E(\hat{C}_{i,I}) = E(C_{i,I})$, $2 \leq i \leq I$.

The proof of this Proposition is postponed to the Appendix A.4.

Remark 4.1.

- If we set $\gamma_{i,j} = \delta_{i,j} = w_{i,j} C_{i,j}^\alpha$, for $\alpha \in \{0,1,2\}$, in (4.2) and (4.4) we get the assumptions of MCL model from Mack (1999) (see also Mack (1993), Mack (1994) and Saito 2009).
- If we put $\gamma_{i,j} = \delta_{i,j} = w_{i,j} C_{i,j}^{\alpha_j}$, for $\alpha_j \in \mathbb{R}$, in (4.2) and (4.4) we get the stochastic chain-ladder model from Murphy, Bardis and Majidi (2012).

5. Main results

5.1. Single accident years

Result 5.1 (Conditional MSEP estimator for a single accident year).

$$\widehat{mse}_{\hat{C}_{i,I}|D_i}(C_{i,I}) = (\hat{C}_{i,I})^2 \cdot (\hat{\Gamma}_{i,I} + \hat{\Delta}_{i,I}), \quad (5.1)$$

where

$$\hat{\Gamma}_{i,I} = \sum_{k=I-i+1}^{J-1} \frac{\hat{\sigma}_k^2 / (\hat{f}_k)^2}{\hat{\delta}_{i,k}}, \quad (5.2)$$

$$\hat{\Delta}_{i,I} = \sum_{k=I-i+1}^{J-1} \frac{\hat{\sigma}_k^2 / (\hat{f}_k)^2}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} \cdot \sum_{j=1}^{I-k} \frac{(\gamma_{j,k})^2}{\delta_{j,k}} \cdot \mathbf{1}_{\{\delta_{j,k} \neq 0\}}, \quad (5.3)$$

and $\hat{f}_j$ and $\hat{\sigma}_j^2$ are given in (4.2) and (4.4)–(4.5), respectively.

5.2. Aggregation over prior accident year

Result 5.2 (Conditional MSEP estimator for aggregated years).

$$\begin{aligned} \widehat{mse}_{\sum_{i=1}^I \hat{C}_{i,I}|D_I}\left(\sum_{i=1}^I C_{i,I}\right) &= \sum_{i=2}^I \widehat{mse}_{\hat{C}_{i,I}|D_i}(C_{i,I}) \\ &+ \sum_{i=2}^I \hat{C}_{i,I} \left(\sum_{j=i+1}^I \hat{C}_{j,I}\right) \sum_{k=I-i+1}^{J-1} 2 \frac{\hat{\sigma}_k^2 / (\hat{f}_k)^2}{\left(\sum_{l=1}^{I-k} \gamma_{l,k}\right)^2} \\ &\cdot \sum_{l=1}^{I-k} \frac{(\gamma_{l,k})^2}{\delta_{l,k}} \cdot \mathbf{1}_{\{\delta_{l,k} \neq 0\}}, \end{aligned} \quad (5.4)$$

where $\hat{f}_j$ and $\hat{\sigma}_j^2$ are defined in (4.2) and (4.4)–(4.5), respectively.

6. Applications of GMCL model

In our numerical example in Section 3.5 we have seen that the assumption about the same weights in estimation of parameters σ_k and f_k yields for some methods to an artificial reduction of variability of reserves amounts (refer to Table 2). To overcome this difficulty, we introduced the different weights $\gamma_{i,j}$ and $\delta_{i,j}$ in computation of $\hat{f}_k$ and $\hat{\sigma}_k$, respectively. In the following application we indicate how one can possibly estimate the weights $\gamma_{i,j}$ and $\delta_{i,j}$ and we point out some other interesting applications.

6.1. Method proxy for factors selection

In this section, we examine our general framework $g_{\gamma,j}(C_{i,j}) := w_{i,j}^\gamma C_{i,j}^\alpha$ and $g_{\delta,j} C_{i,j} := w_{i,j}^\delta C_{i,j}^\beta$, which means that $\gamma_{i,j} := w_{i,j}^\gamma C_{i,j}^\alpha$ and $\delta_{i,j} := w_{i,j}^\delta C_{i,j}^\beta$. In this so called *proxy method* we impose using all link ratios $F_{i,j}$ in estimation of parameters σ_k ($w_{i,j}^\delta = 1$, for all i, j). For all five methods presented, we take $\alpha = \beta = 0$. We turn back to our numerical example from Section 3.5 and we evaluate the same estimators for the already-presented five methods with the only difference in weights of σ_k estimation. More precisely:

- (1) **ALL AV:** The $\hat{\sigma}_k^2$ are estimated with $w_{i,j}^\delta = 1$ for all i, j . Parameters $\hat{f}_k$ are computed as a arithmetic average of all individual link ratios $F_{i,j}$. More precisely, we define the weights in the following way: $w_{i,j}^\gamma = 1$ for all i, j .

- (2) **AEHL:** The $\hat{\sigma}_k^2$ are estimated with $w_{ij}^\delta = 1$ for all i, j . Parameters $\hat{f}_k$ are computed as a arithmetic average of all individual link ratios excluding the highest and the lowest values of $F_{i,j}$. More precisely, we define the weights in the following way: for fixed j , $w_{ij}^\gamma = 0$ for i such that $F_{i,j} = F_{(I-j),j}$ and $F_{i,j} = F_{(1),j}$, where $F_{(k),j}$ for $k = 1, \dots, I-j$ denotes the order statistics of $F_{i,j}$. For remaining indices i , for fixed j , we take $w_{ij}^\gamma = 1$.
- (3) **5 Years AV:** The $\hat{\sigma}_k^2$ are estimated with $w_{ij}^\delta = 1$ for all i, j . Parameters $\hat{f}_k$ are computed as an arithmetic average of individual link ratios $F_{i,j}$ from five latest accidents years. More precisely, we define the weights in the following way: $w_{ij}^\gamma = 1$ for $i = I-j, \dots, I-j-4$. For remaining indices i , for fixed j , we take $w_{ij}^\gamma = 0$.
- (4) **3 Years AV:** The $\hat{\sigma}_k^2$ are estimated with $w_{ij}^\delta = 1$ for all i, j . Parameters $\hat{f}_k$ are computed as an arithmetic average of individual link ratios $F_{i,j}$ from three latest accidents years. More precisely, we define the weights in the following way: $w_{ij}^\gamma = 1$ for $i = I-j, \dots, I-j-2$. For remaining indices i , for fixed j , we take $w_{ij}^\gamma = 0$.
- (5) **Median:** The $\hat{\sigma}_k^2$ are estimated with $w_{ij}^\delta = 1$ for all i, j . Parameters $\hat{f}_k$ are computed as an arithmetic average of individual link ratios $F_{i,j}$ in the way to obtain the sample median. More precisely, we put $w_{ij}^\gamma = 1$ or $w_{ij}^\gamma = 0$ in the way that the estimators of the age-to-age factors f_k are given by

$$\hat{f}_k = \text{median}\{F_{i,k}: i \in \{1, \dots, I-k\}\},$$

where *median* denotes the sample median which, for the sample $X_1, \dots, X_n$, is computed by

$$\text{median}\{X_i: i \in \{1, \dots, n\}\} := \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd} \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n+1}{2}\right)}}{2} & \text{otherwise} \end{cases}$$

and $X_{(k)}$ denotes the k th order statistics of the sample $X_1, \dots, X_n$.

In Table 3 we present the estimation of $\hat{R}$ and $MSEP(\hat{R})$ using the five methods described above. In terms of MSEP we see that, in general, we have the values greater than our reference method *ALL AV* from column (1), which stays unchanged compared to Table 2. This is not surprising because by selecting of the development factors we decreased the estimated values of f_k and by using all observations $F_{i,j}$ in $\hat{\sigma}_k$ computation we mechanically increased the dispersion around the values of $\hat{f}_k$. In view of our results from Tables 2 and 3, the proxy method overestimates in general the real MSEP, and can then be treated as its upper bound. However, it can be useful as a tool to perform the sensitivity analysis for testing the impact on the reserve volatility of excluding the specific set of link ratios. Finally, it can be seen as a measure of relative prudence of other approach of measuring the variability of reserves by means of MSEP estimators.

Table 3. Estimators of $\hat{R}$, $\sqrt{MSEP(\hat{R})}$ and $CV(\hat{R})$

Item/method	alpha = 0				
	ALL AV (1)	AEHL (2)	5 Years AV (3)	3 Years AV (4)	Median (5)
$\hat{R}$	93 643	65 868	75 886	68 645	54 059
$MSEP(\hat{R})^{1/2}$	92 549	88 105	101 643	113 904	105 786
$CV(\hat{R})$	99%	134%	134%	166%	196%
	(1)/(1)	(2)/(1)	(3)/(1)	(4)/(1)	(5)/(1)
$\hat{R}$ (%)	100%	70%	81%	73%	58%
$MSEP(\hat{R})^{1/2}$ (%)	100%	95%	110%	123%	114%

6.2. Robust estimation in GMCL model

From the previous two numerical examples (MCL vs. GMCL results), we observe that, in general, the first approach underestimates and second overestimates the MSE of claims reserves (see Tables 2 and 3). In this section we present an intermediate solution for our general problem that allows us to evaluate the estimation of MSE of reserves in case of development factor selection. This go-between solution is based on the *robust statistics* in estimation of model parameters f_k and σ_k . The term *robust statistics* is meant in the sense of Huber and Ronchetti (2009).

As already mentioned, the assumption GMCL.2 about the conditional variance of $F_{i,j}$ allows us to estimate the factors f_k in the framework of linear regression obtained by the means of weighted least squares procedure (see Murphy, Bardis, and Majidi 2012 and the references therein). Although these estimators are easy to compute and have excellent theoretical properties (see Proposition 4.1), they rely on quite strict assumptions, and their violation may lead to useless results.

One possible solution to overcome this difficulty is to use robust estimation techniques. The idea of robust statistics is to account for certain deviations from idealized model assumptions. Typically, robust methods reduce the influence of outlying observation on the estimator.

We take the following assumption in our general framework of GMCL model: $g_{\gamma_j}(t) := t^{\alpha_j}$ and $g_{\delta_j}(t) := t^{\beta_j}$, with $\alpha_j, \beta_j \in \mathbb{R}$ to be estimated. This means that $\gamma_{ij} := C_{ij}^{\alpha_j}$ and $\delta_{ij} := C_{ij}^{\beta_j}$.

The following algorithm shows how one can estimate the parameters α_j for $j = 1, \dots, I-1$ and β_k for $k = 1, \dots, I-2$. As can be seen, the presented method is based on a similar principle to the well known moment estimation method from point estimation theory.

6.2.1. Algorithm for fitting α_k and β_k parameters

Step 1. We select the robust estimators for f_k and its variance. We denote these estimators by $\tilde{f}_k$

and $\tilde{Var}(\tilde{f}_k)$ respectively. These two quantities can be derived by numerous techniques described in the literature, such as: M-estimation, L^p estimation, etc. (Huber and Ronchetti 2009) or trimmed mean (Jeng 2010).

Step 2. For every $k = 1, \dots, I-1$, we find α_k by solving the following equation $\hat{f}_k = \tilde{f}_k$, where $\hat{f}_k$ is given in equation (4.2), namely

$$\frac{\sum_{i=1}^{I-k} C_{i,k}^{\alpha_k} F_{i,k}}{\sum_{i=1}^{I-k} C_{i,k}^{\alpha_k}} = \tilde{f}_k. \quad (6.1)$$

The procedure to select the *consistent* α_k together with the problem of existence of solution of equation (6.1) is treated in Murphy, Bardis, and Majidi (2012) (see Lemma 1 and the comments that follow it).

Step 3. For every $k = 1, \dots, I-2$, we find β_k by solving following equation: $\hat{Var}(\hat{f}_k) = \tilde{Var}(\tilde{f}_k)$, where $\hat{Var}(\hat{f}_k)$ is given in (A.8), namely,

$$\frac{1}{I-k-1} \sum_{i=1}^{I-k} C_{i,k}^{\beta_k} (F_{i,k} - \hat{f}_k)^2 \cdot \frac{\sum_{j=1}^{I-k} \gamma_{j,k}^2 C_{i,k}^{\beta_k}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k} \right)^2} = \tilde{Var}(\tilde{f}_k) \quad (6.2)$$

The parameters $\hat{f}_k$ are given in equation (4.2) with α_k estimated in **Step 2**. For $k = I-1$, since only one observation is available in our data, β_{I-1} need to be estimated by other approaches. The limits for the values of $\tilde{Var}(\tilde{f}_k)$ for which the solution of equation (6.2) exists are presented in Appendix E.

6.2.2. Numerical example

In the present example we consider only the *median* method already presented in previous numerical applications. We concentrate on that particular method because our goal is not to present the extensive case study but rather to illustrate the general principle and the main steps of this application. Observe that the method *AEHL* can be treated by the theory of robust

Table 4. Estimation of $\tilde{f}_k$ and $s.e(\tilde{f}_k)$ using LAD technique corresponding to sample median method

k	1	2	3	4	5	6	7	8	9
$\tilde{f}_k$	4,2597	1,5992	1,1635	1,1657	1,1318	1,0335	1,0333	1,0180	1,0092
$s.e(\tilde{f}_k)$	1,7974	0,1686	0,1411	0,0270	0,0472	0,0251	0,0065	0,0129	—
$CV(\tilde{f}_k)$	42,2%	10,5%	12,1%	2,3%	4,2%	2,4%	0,6%	1,3%	—

Table 5. Estimation of parameters $\hat{\alpha}_j$ and $\hat{\beta}_j$

j	1	2	3	4	5	6	7	8	9
$\hat{\alpha}_j$	0,5204	1,4073	0	1,5852	-0,3835	0	1,0022	0	0,0000
$\hat{\beta}_j$	0,6605	0,3120	0	0,7207	1,9501	0	-2,7733	0	—

estimation by means of *trimmed mean* estimators (Jeng 2010).

To apply the above fitting algorithm for sample *median* method, we use the LAD (least absolute deviation) estimation procedure. The theoretical framework of LAD is presented in Appendix D. The values of $\tilde{f}_k$ are given by computing the sample median from $F_{i,k}$ as described in Section 3.5. The standard errors of $\tilde{f}_k$ are obtained via bootstrap techniques. In Table 4 we present the numerical values of $\tilde{f}_k$, $s.e(\tilde{f}_k) := \tilde{Var}(\tilde{f}_k)^{1/2}$ and $CV(\tilde{f}_k) := s.e(\tilde{f}_k)/\tilde{f}_k$. Note that the last value of $s.e(\tilde{f}_k)$ cannot be estimated from the data for the reasons discussed in the case of σ_{l-1} estimation in Section 3.2 (only one observation available). This why we do not fit the parameter β_{l-1} via equation (6.2), but we put $\beta_{l-1} := \alpha_{l-1}$.

The standard deviations of $\tilde{f}_k$ from Table 4 were obtained by using the *rq* function integrated in free R software. Note that the value of $\tilde{f}_k$ given by this function are slightly different from those presented in Section 3.2. This is probably due to the optimization algorithm that is used in R. Given that these differences are insignificant, we decided to present in Table 4 the same numerical values of the estimators $\tilde{f}_k$ as given in Section 3.2. The corresponding R code is available on request from the author.

As mentioned in the algorithm and in Appendix E, the solutions of (6.1) and (6.2) are not always available. For instance, in our example, there is no solution of equation (6.1) for $k = 3, 4$ and no solution of equation (6.2) for $k = 8$. This means that for $k = 3$,

$k = 4$, and $k = 8$ the parameters α_k and β_k need to be specified in different way. This could be done using any other approach that is being judged appropriate by the actuary performing estimation. In our case, we put $\alpha_3 = \beta_3 = \alpha_6 = \beta_6 = \alpha_8 = \beta_8 = 0$ to have from one hand the optimal properties (see Proposition 4.1) but also to be consistent with our choice of $\alpha = 0$ in our two previous numerical applications (see Sections 3.5 and 6.1).

The estimation of parameters α_j and β_j are stated in Table 5. The values of $\hat{\alpha}_j$ and $\hat{\beta}_j$ for which we arbitrarily put 0 are indicated with bold font characters.

The MSE and claims reserves amount estimators are stated in Table 6. Observe that the robust estimation is a good compromise between the method with the same weights (see Section 3.5) and the method where we use all link ratios in σ_k estimation (see *proxy method* in Section 6.1).

6.3. Validation of results from reserving softwares

The next interesting and extremely important application of our GMCL model is the possibility of

Table 6. Median method with robust estimation

Item/method	Median		
	MCL	Robust	Proxy
$\hat{R}$	54 059	63 165	54 059
$MSEP(\hat{R})^{1/2}$	14 786	40 312	105 786
$CV(\hat{R})$	27%	64%	196%

validating the results from industry reserving software. The stochastic chain-ladder type methods are used to evaluate the economic risk capital required by Solvency II for so-called *reserve risk*. In fact, this capital requirement for reserve risk is computed as the 99.5th percentile (value at risk) of run-off result distribution (profit/loss on reserves over one year). This means that Solvency II defines the reserve risk in one-year time horizon, which is different from the standard approach considering the distribution of the ultimate cost of claims.

However, one of the methods to derive the one-year reserve risk is based on simple scaling of ultimate view. This technique is based on using the results of Merz and Wüthrich (2008), which is currently a popular methodology throughout the market and taken from the latest technical literature on this topic.

The empirical loss distribution in ultimate view is often derived by using the bootstrap techniques and Monte Carlo simulations. The first technique is used to evaluate the estimation error and the second to approximate the process variance. This kind of bootstrap approach is also available in ResQ software, which is used worldwide within the P&C insurance market. The question is how to validate the results from bootstrap method provided by reserving tools such as ResQ. One of the possible solutions is to compare the estimation of the first two moments of loss distribution from bootstrapping (based on simulations) with the estimators of reserves and MSEP of reserves obtained by the explicit formulas. For the sake of simplicity, we assume that there is no factors selection (all weights $w_{i,j}$ are fixed to 1). We use the RAA run-off triangle and we present the numerical results in Table 7. For all bootstrapping

results we used 100 000 simulations. We begin our analysis with the classic chain-ladder method in which the estimators of f_k are the all volume weighted average and are consistent with Mack (1993). More precisely, with the hypothesis of the MCL method with $\alpha = 1$, we compare the estimate of MSEP obtained by these two techniques: bootstrap from ResQ and explicit formula given in MCL approach. The corresponding numerical values are respectively: 27 150 (see ResQ(Boot) (3) in Table 7) and 26 909 (see ResQ(MCL) (4) in Table 7). We observe a good convergence for bootstrap (the relative error is less than 1%). We consider now the different estimator of f_k computed as a simple arithmetic average of individual link ratios $F_{i,j}$. This is equivalent to taking $\alpha = 0$ in the MCL framework. In that case, we observe that the estimates of MSEP for both methods become divergent: 75 656 (see ResQ(Mack) (2) in Table 7) and 58 475 (see ResQ(Boot) (1) in Table 7). This is due to the fact that the ResQ(Mack) method is obtained by approximation based on the MCL formula with $\alpha = 1$. In fact, according to the technical documentation, the ResQ estimates of parameters f_k and σ_k in the bootstrap approach are of the form (up to multiplicative constant for bias reduction):

$$\hat{f}_k = \frac{1}{I-k} \sum_{i=1}^{I-k} F_{i,k} \text{ and } \sigma_k^2 = \frac{1}{I-k-1} \sum_{i=1}^{I-k} C_{i,k} (F_{i,k} - \hat{f}_k)^2.$$

It is easily seen that these estimators are consistent with our general approach with $\alpha = 0$ and $\beta = 1$ (see (4.2) and (4.4) in Section 4). The MSEP estimator is equal to 59 065 (see GMCL (5) in Table 7). This shows that GMCL method allows one to validate the results and detect the incoherences. Effectively, the choice of estimators in ResQ for the case $\alpha = 0$ is not optimal in sense of Proposition 4.1. It remains

Table 7. Comparison of ResQ estimators of $\hat{R}$ and $MSEP(\hat{R})$ with MCL and GMCL models

Item/method	alpha = 0		alpha = 1		alpha = 0, beta = 1
	ResQ(Boot) (1)	ResQ(Mack) (2)	ResQ(Boot) (3)	MCL (4)	GMCL (5)
$\hat{R}$	93 630	93 643	52 204	52 135	93 643
$MSEP(\hat{R})^{1/2}$	58 475	75 656	27 150	26 909	59 065

unknown whether this is deliberate or whether this is just a proxy approach that was judged correct.

In regards to the approximation ResQ(Mack) (2), this shows that in construction of the proxy methods we cannot just take the MSEP formula for $\alpha = 1$ as a starting point. Indeed, the MSEP formula changes if we modify the estimates of f_k because the variance of f_k is not the same, so it is not enough to plug in the new estimators of $C_{i,l}$, f_k and σ_k in the MSEP formula (5.4) with $\alpha = \beta = 1$. This lack of understanding of this principle could be a reason of taking the no optimal hypothesis in bootstrap ResQ(Boot) (1) method.

Finally, we observe that the results of ResQ(Boot) (1) method validate our explicit formula for estimation of MSEP of claims reserves in the framework of our GMCL model.

7. Conclusion

In this paper we presented a general flexible tool for stochastic loss reserving and its variability. We developed our GMCL model to quantify the variability of reserves in the context of selecting development factors in the framework of the stochastic chain-ladder method.

We provided the theoretical and flexible background which covers some practices of actuaries and industrial providers of reserving softwares.

Finally, we showed the way of bridging the chain ladder model and the robust estimation techniques. Our results can be applied in other approaches based on chain-ladder framework like: multivariate chain-ladder, univariate and multivariate Bayesian chain-ladder, etc. One can derive the similar results in the context of one-year reserve risk for Solvency II purposes. This topic will be treated in our forthcoming paper. Some partial results can be found in Sloma (2014) and Sloma (2011).

Acknowledgments

The author want to thank the reviewers for their helpful comments and suggestions.

Appendix: Mathematical Proofs

We present here the proofs of our main results. Most of them are derived by simple rewriting the techniques applied in Mack (1993), Mack (1994).

A.1. Proof of Result 5.1

Due to the general rule $E(X - c)^2 = \text{Var}(X) + (EX - c)^2$ for any scalar c we have

$$\begin{aligned} \text{mse}_{\hat{C}_{il}|D_I}(C_{il}) &= E\left[\left(\hat{C}_{il} - C_{il}\right)^2 \middle| D_I\right] \\ &= \text{Var}(C_{il}|D_I) + \left(E(C_{il}|D_I) - \hat{C}_{il}\right)^2. \end{aligned} \quad (\text{A.1})$$

To estimate $\text{Var}(C_{i,l}|D_I)$ we use the following

Lemma 9.1. For $i = 2, \dots, I$, we have,

$$\text{Var}(C_{i,l}|D_I) = \sum_{l=I+1-i}^{I-1} E\left[\frac{C_{i,l}^2}{\delta_{i,l}} \middle| D_I\right] \sigma_l^2 \prod_{k=l+1}^{I-1} f_k^2. \quad (\text{A.2})$$

The proof of Lemma A.1 is postponed to Appendix A.5.

Note that the estimation of $E\left[\frac{C_{i,l}^2}{\delta_{i,l}} \middle| D_I\right]$ from equation (A.2) is a crucial part of this proof. We choose to estimate this term by $\frac{\hat{C}_{i,l}^2}{\hat{\delta}_{i,l}}$. This is due to the obvious

observation that $\frac{C_{i,l}^2}{\delta_{i,l}}$ is an unbiased estimate of

$E\left[\frac{C_{i,l}^2}{\delta_{i,l}}\right]$ and from the basic property of conditional expectation, namely: $E\left[E\left[\frac{C_{i,l}^2}{\delta_{i,l}} \middle| D_I\right]\right] = E\left[\frac{C_{i,l}^2}{\delta_{i,l}}\right]$.

It is worth noting here that, in the case where $\delta_{i,l} := C_{i,l}^\alpha$ with $\alpha \in \mathbb{R}$, in Saito (2009), the author used the same technique of estimation without giving any justification or reason for that (see proof of Lemma 4 and Estimate 8). Similarly, in the case where $\delta_{i,l} := w_{i,l} \cdot C_{i,l}^\alpha$ with $\alpha \in \{0, 1, 2\}$ and $w_{i,l} \in [0, 1]$, we find the same estimator in Mack (1999). More precisely, the author claims (without proving) that

$\sum_{l=I+1-i}^{I-1} \text{Var}(C_{i,l+1}|D_l) \prod_{k=l+1}^{I-1} f_k^2$ can be estimated via the quantity $\hat{C}_{i,l} \sum_{l=I+1-i}^{I-1} (s.e.(F_{i,l}))^2 / \hat{f}_l^2$, where $(s.e.(F_{i,l}))^2$ is an estimate of $\text{Var}(F_{i,l}|C_{i,1}, \dots, C_{i,l})$. Indeed, this is achieved if we estimate $E\left[\frac{C_{i,l}^2}{w_{i,l} \cdot C_{i,l}^\alpha} \middle| D_l\right]$ by $\frac{\hat{C}_{i,l}^2}{w_{i,l} \cdot \hat{C}_{i,l}^\alpha}$.

However, in Murphy, Bardis, and Majidi (2012), the authors used different approach based on normal approximation.

Note that in the Section 6.3 we obtained that the above estimator of $\text{Var}(C_{i,l+1}|D_l)$ is consistent with that provided by bootstrap technique from ResQ software. This is shown for the particular run-off triangle and the assumption that $C_{i,j}$ are gamma-distributed random variables. It would be interesting to perform the extensive simulation study in order to examine the exactitude of this estimate with other data and probability distributions.

We apply now Lemma A.1 with $\hat{E}\left[\frac{C_{i,l}^2}{\delta_{i,l}} \middle| D_l\right] = \frac{\hat{C}_{i,l}^2}{\hat{\delta}_{i,l}}$ and by replacing the unknown parameters f_k et σ_k^2 with their estimators $\hat{f}_k$ and $\hat{\sigma}_k^2$. Together with the equality $\hat{C}_{i,l} = C_{i,l+1} \prod_{k=l+1-i}^{I-1} \hat{f}_k$ (see Proposition 4.1 (v)) we conclude

$$\begin{aligned} \text{Var}(C_{i,l}|D_l) &= \sum_{l=I+1-i}^{I-1} E\left[\frac{C_{i,l}^2}{\delta_{i,l}} \middle| D_l\right] \sigma_l^2 \prod_{k=l+1}^{I-1} f_k^2 \\ &= \sum_{l=I+1-i}^{I-1} \frac{\hat{C}_{i,l}^2}{\hat{\delta}_{i,l}} \hat{\sigma}_l^2 \prod_{k=l+1}^{I-1} \hat{f}_k^2 \\ &= \sum_{l=I+1-i}^{I-1} \frac{\hat{\sigma}_l^2}{\hat{\delta}_{i,l}} C_{i,l+1-i}^2 \prod_{k=l+1-i}^{I-1} \hat{f}_k^2 \cdot \prod_{k=l+1}^{I-1} \hat{f}_k^2 \\ &= C_{i,l+1-i}^2 \sum_{l=I+1-i}^{I-1} \frac{\hat{\sigma}_l^2 / \hat{f}_l^2}{\hat{\delta}_{i,l}} \prod_{k=l+1-i}^{I-1} \hat{f}_k^2 \\ &= C_{i,l+1-i}^2 \cdot \prod_{k=l+1-i}^{I-1} \hat{f}_k^2 \sum_{l=I+1-i}^{I-1} \frac{\hat{\sigma}_l^2 / \hat{f}_l^2}{\hat{\delta}_{i,l}} \\ &= \hat{C}_{i,l}^2 \sum_{l=I+1-i}^{I-1} \frac{\hat{\sigma}_l^2 / \hat{f}_l^2}{\hat{\delta}_{i,l}}. \end{aligned} \quad (\text{A.3})$$

We now turn to the second summand of the expression (A.1). Because of Proposition 4.1 (iv) and (v) we have,

$$\begin{aligned} &(E(C_{i,l}|D_l) - \hat{C}_{i,l})^2 \\ &= C_{i,l+1-i}^2 (f_{l+1-i} \cdot \dots \cdot f_{l-1} - \hat{f}_{l+1-i} \cdot \dots \cdot \hat{f}_{l-1})^2. \end{aligned} \quad (\text{A.4})$$

As can be easily seen, this expression cannot be estimated by replacing f_k with $\hat{f}_k$. In order to estimate the right hand side of (A.4) we use the same approach as in Mack (1993), Mack (1994). Saito (2009) followed the same technique of estimation. However, in Murphy, Bardis, and Majidi (2012) we can find different approach which was also presented in Buchwalder et al. (2006a). It is worth noting that in the paper of Mack, Quarg, and Braun (2006) the authors criticised the approaches of Buchwalder et al. (2006a) and showed that the estimate of estimation error from Mack (1993) is hard to be improved (see also Buchwalder et al. 2006b). As the answer for the criticism of Mack on article of Buchwalder et al. (2006a), the authors provided the bounds for estimation error and claimed that the Mack estimator, in some particular cases, is closed to these bounds (see Wüthrich, Merz, and Bühlmann 2008). This should be confirmed by performing the extensive simulation study to quantify the different approaches of error estimation in stochastic Chain-Ladder framework.

We define,

$$\begin{aligned} F &= f_{l+1-i} \cdot \dots \cdot f_{l-1} - \hat{f}_{l+1-i} \cdot \dots \cdot \hat{f}_{l-1} \\ &= S_{l+1-i} + \dots + S_{l-1}, \end{aligned} \quad (\text{A.5})$$

with

$$\begin{aligned} S_k &= \hat{f}_{l+1-i} \cdot \dots \cdot \hat{f}_{k-1} f_k f_{k+1} \cdot \dots \cdot f_{l-1} \\ &\quad - \hat{f}_{l+1-i} \cdot \dots \cdot \hat{f}_{k-1} \hat{f}_k \hat{f}_{k+1} \cdot \dots \cdot \hat{f}_{l-1} \\ &= \hat{f}_{l+1-i} \cdot \dots \cdot \hat{f}_{k-1} (f_k - \hat{f}_k) f_{k+1} \cdot \dots \cdot f_{l-1}. \end{aligned} \quad (\text{A.6})$$

This yields

$$\begin{aligned} F^2 &= (S_{l+1-i} + \dots + S_{l-1})^2 \\ &= \sum_{k=l+1-i}^{I-1} S_k^2 + 2 \sum_{k=l+1-i}^{I-1} \sum_{j < k}^{I-1} S_j S_k. \end{aligned} \quad (\text{A.7})$$

We estimate F^2 using the following

Proposition A.1 (Estimate of F^2) Let define, for $1 \leq k \leq I$, the set of observed C_{ij} up to development year k , namely

$$B_k = \{C_{ij} : i + j \leq I + 1, k \leq j\} \subset D_I.$$

Then, we can estimate F^2 by

$$\widehat{F}^2 = \prod_{l=I+1-i}^{I-1} \hat{f}_l^2 \sum_{k=I+1-i}^{I-1} \frac{\text{Var}(\hat{f}_k | B_k)}{\hat{f}_k^2}.$$

The proof of Proposition A.1 is postponed to Appendix A.5.

It remains to determine the estimate of $\text{Var}(\hat{f}_k | B_k)$. We use the following

Proposition A.2 We assume (4.3). We have

$$\text{Var}(\hat{f}_k | B_k) = \sigma_k^2 \frac{\sum_{j=1}^{I-k} \frac{\gamma_{j,k}^2}{\delta_{j,k}} \cdot \mathbf{1}_{\{\delta_{j,k} \neq 0\}}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k} \right)^2}. \quad (\text{A.8})$$

The proof of Proposition A.2 is postponed to Appendix A.5.

Finally, using (A.4) and Proposition A.2 we estimate $E(C_{i,l} | D_l) - \hat{C}_{i,l}$ by

$$\begin{aligned} & C_{i,I+1-i}^2 \hat{f}_{I+1-i}^2 \cdots \hat{f}_{I-1}^2 \sum_{k=I+1-i}^{I-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \frac{\sum_{j=1}^{I-k} \frac{\gamma_{j,k}^2}{\delta_{j,k}} \cdot \mathbf{1}_{\{\delta_{j,k} \neq 0\}}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k} \right)^2} \\ &= \hat{C}_{i,l}^2 \sum_{k=I+1-i}^{I-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \frac{\sum_{j=1}^{I-k} \frac{\gamma_{j,k}^2}{\delta_{j,k}} \cdot \mathbf{1}_{\{\delta_{j,k} \neq 0\}}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k} \right)^2}. \end{aligned}$$

This completes the proof of Result 5.1. ■

A.2. Proof of result 5.2 (overall standard error)

Following the definition in (2.4), we have

$$\begin{aligned} msep_{\sum_{i=1}^I \hat{C}_{i,l} | D_l} \left(\sum_{i=1}^I C_{i,l} \right) &= E \left[\left(\sum_{i=1}^I \hat{C}_{i,l} - \sum_{i=1}^I C_{i,l} \right)^2 \middle| D_l \right] \\ &= \text{Var} \left(\sum_{i=1}^I C_{i,l} | D_l \right) + \left(E \left(\sum_{i=1}^I C_{i,l} | D_l \right) - \sum_{i=1}^I \hat{C}_{i,l} \right)^2 \end{aligned}$$

The independence of accident years yields

$$\text{Var} \left(\sum_{i=1}^I C_{i,l} | D_l \right) = \sum_{i=1}^I \text{Var} (C_{i,l} | D_l).$$

where each term of the sum has already been calculated in the proof of the Result 5.1.

Furthermore

$$\begin{aligned} & \left(E \left(\sum_{i=1}^I C_{i,l} | D_l \right) - \sum_{i=1}^I \hat{C}_{i,l} \right)^2 \\ &= \left(\sum_{i=1}^I (E(C_{i,l} | D_l) - \hat{C}_{i,l}) \right)^2 \\ &= \sum_{i,j} (E(C_{i,l} | D_l) - \hat{C}_{i,l})(E(C_{j,l} | D_l) - \hat{C}_{j,l}) \end{aligned}$$

Taking together

$$\begin{aligned} \widehat{msep}_{\sum_{i=1}^I \hat{C}_{i,l} | D_l} \left(\sum_{i=1}^I C_{i,l} \right) &= \sum_{i=2}^I \widehat{msep}_{\hat{C}_{i,l} | D_l} (C_{i,l}) \\ &\quad + \sum_{2 \leq i \leq j \leq I} 2 \cdot C_{i,I+1-i} C_{j,I+1-j} F_i F_j, \end{aligned}$$

with

$$F_i = f_{I+1-i} \cdots f_{I-1} - \hat{f}_{I+1-i} \cdots \hat{f}_{I-1} = \sum_{k=I+1-i}^{I-1} S_k^i,$$

where

$$S_k^i = \hat{f}_{I+1-i} \cdots \hat{f}_{k-1} (f_k - \hat{f}_k) f_{k+1} \cdots f_{I-1}.$$

We can determine the estimator of $F_i F_j$ in the analogous way as for F^2 .

Proposition A.3. *We have*

$$\widehat{F_i F_j} = \sum_{k=I+1-i}^{I-1} \frac{\text{Var}(\hat{f}_k | B_k)}{\hat{f}_k^2} (\hat{f}_{I+1-i} \cdots \hat{f}_{I-1}) \cdot (\hat{f}_{I+1-j} \cdots \hat{f}_{I-1}).$$

We finally conclude, from Proposition A.3

$$\begin{aligned} & \sum_{2 \leq i < j \leq I} 2 \cdot C_{i,I+1-i} C_{j,I+1-j} \widehat{F_i F_j} \\ &= \sum_{i=2}^I \hat{C}_{i,I} \sum_{j=i+1}^I \hat{C}_{j,I} \sum_{k=I-i+1}^{I-1} 2 \frac{\hat{\sigma}_k^2 / \hat{f}_k^2}{\left(\sum_{l=1}^{I-k} \gamma_{l,k} \right)^2} \cdot \sum_{l=1}^{I-k} \frac{\gamma_{l,k} \cdot \mathbf{1}_{\{\delta_{l,k} \neq 0\}}}{\delta_{l,k}}. \end{aligned}$$

A.3. Proof of Proposition 3.1

- i. See Theorem 2 p. 215 in Mack (1993).
- ii. See discussion on p. 112, Corollary on p. 141 and Appendix B on p. 140 in Mack (1994).
- iii. See Appendix E on p. 151 in Mack (1994).
- iv. See Theorem 1 p. 215 and discussion after the proof of Theorem 2 on page 216 in Mack (1993).
- v. see Appendix C p. 142 in Mack (1994):

A.4. Proof of Proposition 4.1

- (i), (iv) and (v), see proofs of (i), (iv) and (v) respectively in Proposition 3.1.
- (ii). The first part of the statement regarding to the minimal variance of parameters f_k can be easily derived from the proof of (ii) in Proposition 3.1. The rest of the proof is easily seen from the Proposition A.2.
- (iii). Without loss of generality and to avoid the complexity of notation we present the proof for $I_k = I - k$ (for each k , all weights $\delta_{i,k}$ are different from 0).

We have, for $1 \leq k \leq I - 2$,

$$\begin{aligned} (I - k - 1) \cdot \hat{\sigma}_k^2 &= \sum_{i=1}^{I-k} \delta_{i,k} (F_{i,k} - \hat{f}_k)^2 \\ &= \sum_{i=1}^{I-k} \delta_{i,k} F_{i,k}^2 - 2 \sum_{i=1}^{I-k} \delta_{i,k} F_{i,k} \cdot \hat{f}_k + \sum_{i=1}^{I-k} \delta_{i,k} \hat{f}_k^2. \end{aligned}$$

Since $\delta_{i,k}$ are $\sigma(C_{i,k})$ measurable, we have

$$\begin{aligned} E((I - k - 1) \cdot \hat{\sigma}_k^2 | B_k) &= \sum_{i=1}^{I-k} \delta_{i,k} E(F_{i,k}^2 | B_k) \\ &\quad - 2 \sum_{i=1}^{I-k} \delta_{i,k} E(F_{i,k} \cdot \hat{f}_k | B_k) \\ &\quad + \sum_{i=1}^{I-k} \delta_{i,k} E(\hat{f}_k^2 | B_k). \end{aligned}$$

In the following derivation we use $\sigma(C_{i,k})$ -measurability of $\gamma_{i,k}$ and definition of $\hat{f}_k$ from (4.2). Furthermore, the assumption GMCL.3 implies that $F_{i,k}$ and $F_{j,k}$ are independent for $i \neq j$. From assumption GMCL.1 and GMCL.2 we easily see that $E(F_{i,k}^2 | B_k) = \frac{\sigma_k^2}{\delta_{i,k}} + f_k^2$. Taking together,

$$\begin{aligned} E(F_{i,k} \cdot \hat{f}_k | B_k) &= \frac{1}{\sum_{l=1}^{I-k} \gamma_{l,k}} \left(\sum_{j=1}^{I-k} \gamma_{j,k} \cdot E(F_{i,k} \cdot F_{j,k} | B_k) \right) \\ &= \frac{1}{\sum_{l=1}^{I-k} \gamma_{l,k}} \left(\gamma_{i,k} \cdot E(F_{i,k}^2 | B_k) + \sum_{j \neq i}^{I-k} \gamma_{j,k} \cdot E(F_{i,k} | B_k) \cdot E(F_{j,k} | B_k) \right) \\ &= \frac{1}{\sum_{l=1}^{I-k} \gamma_{l,k}} \left(\gamma_{i,k} \cdot \left(\frac{\sigma_k^2}{\delta_{i,k}} + f_k^2 \right) + \sum_{j \neq i}^{I-k} \gamma_{j,k} f_k^2 \right) \\ &= \frac{1}{\sum_{l=1}^{I-k} \gamma_{l,k}} \left(\gamma_{i,k} \cdot \frac{\sigma_k^2}{\delta_{i,k}} + f_k^2 \sum_{j=1}^{I-k} \gamma_{j,k} \right) \\ &= \sigma_k^2 \frac{\gamma_{i,k}}{\sum_{l=1}^{I-k} \gamma_{l,k} \delta_{l,k}} + f_k^2. \end{aligned} \tag{A.9}$$

From Proposition A.2

$$\begin{aligned} E(\hat{f}_k^2 | B_k) &= \text{Var}(\hat{f}_k | B_k) + (E(\hat{f}_k | B_k))^2 \\ &= \sigma_k^2 \frac{\sum_{j=1}^{I-k} \gamma_{j,k}^2}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} + f_k^2. \end{aligned}$$

Taking together we have

$$\begin{aligned} E((I-k-1) \cdot \hat{\sigma}_k^2 | B_k) &= \sum_{i=1}^{I-k} \delta_{i,k} E(F_{i,k}^2 | B_k) - 2 \sum_{i=1}^{I-k} \delta_{i,k} E(F_{i,k} \cdot \hat{f}_k | B_k) \\ &\quad + \sum_{i=1}^{I-k} \delta_{i,k} E(\hat{f}_k^2 | B_k) \\ &= \sum_{i=1}^{I-k} \delta_{i,k} \left(\frac{\sigma_k^2}{\delta_{i,k}} + f_k^2 \right) - 2 \sum_{i=1}^{I-k} \delta_{i,k} \left(\sigma_k^2 \frac{\gamma_{i,k}}{\sum_{j=1}^{I-k} \gamma_{j,k}} + f_k^2 \right) \\ &\quad + \sum_{i=1}^{I-k} \delta_{i,k} \left(\sigma_k^2 \frac{\sum_{j=1}^{I-k} \gamma_{j,k}^2}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} + f_k^2 \right) \\ &= (I-k) \sigma_k^2 + f_k^2 \sum_{i=1}^{I-k} \delta_{i,k} - 2 \sigma_k^2 \sum_{i=1}^{I-k} \delta_{i,k} \frac{\gamma_{i,k}}{\sum_{j=1}^{I-k} \gamma_{j,k}} \\ &\quad + \sigma_k^2 - \frac{\sum_{i=1}^{I-k} \delta_{i,k} \sum_{j=1}^{I-k} \gamma_{j,k}^2}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} + f_k^2 \sum_{i=1}^{I-k} \delta_{i,k} \\ &= (I-k-1) \sigma_k^2 + \sigma_k^2 \left[\frac{\sum_{i=1}^{I-k} \delta_{i,k} \sum_{j=1}^{I-k} \gamma_{j,k}^2}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} - 1 \right]. \quad (\text{A.10}) \end{aligned}$$

Finally

$$\begin{aligned} E(\hat{\sigma}_k^2 - \sigma_k^2) &= E[E((\hat{\sigma}_k^2 - \sigma_k^2) | B_k)] \\ &= \frac{\sigma_k^2}{I-k-1} E \left[\frac{\sum_{i=1}^{I-k} \delta_{i,k} \sum_{j=1}^{I-k} \gamma_{j,k}^2}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} - 1 \right]. \end{aligned}$$

A.5. Proofs of auxiliary results

A.5.1. Proof of Lemma 9.1

Let define, for $1 \leq i \leq I$ and $1 \leq j \leq I$, the set of observed data $C_{i,j}$ for accident year i and up to development year j , namely

$$A_{i,j} = \{C_{i,k} : 1 \leq k \leq j\}.$$

For $l = I+1-i, \dots, I-1$,

$$\begin{aligned} \text{Var}(C_{i,l} | D_l) &= \text{Var}(C_{i,l+1} | A_{i,l+1-i}) \\ &\quad E[\text{Var}(C_{i,l+1} | A_{i,l+1-i}) | A_{i,l+1-i}] \\ &\quad + \text{Var}[E(C_{i,l+1} | A_{i,l+1-i}) | A_{i,l+1-i}] \\ &= E \left[\sigma_l^2 \frac{C_{i,l}^2}{\delta_{i,l}} | A_{i,l+1-i} \right] + \text{Var}[f_l C_{i,l} | A_{i,l+1-i}] \\ &= \sigma_l^2 E \left[\frac{C_{i,l}^2}{\delta_{i,l}} | A_{i,l+1-i} \right] + f_l^2 \text{Var}[C_{i,l} | A_{i,l+1-i}] \\ &= \sigma_l^2 E \left[\frac{C_{i,l}^2}{\delta_{i,l}} | D_l \right] + f_l^2 \text{Var}[C_{i,l} | D_l]. \quad (\text{A.11}) \end{aligned}$$

We multiply the both sides by $\prod_{k=l+1}^{I-1} f_k^2$ with the convention that an empty product equals 1. Taking the sum over $l = I+1-i, \dots, I-1$, we obtain

$$\begin{aligned} &\sum_{l=I+1-i}^{I-1} \text{Var}(C_{i,l+1} | D_l) \prod_{k=l+1}^{I-1} f_k^2 \\ &= \sum_{l=I+1-i}^{I-1} \sigma_l^2 E \left[\frac{C_{i,l}^2}{\delta_{i,l}} | D_l \right] \prod_{k=l+1}^{I-1} f_k^2 \\ &\quad + \sum_{l=I+1-i}^{I-1} \text{Var}[C_{i,l} | D_l] f_l^2 \prod_{k=l}^{I-1} f_k^2 \text{Var}(C_{i,l} | D_l) \\ &\quad + \sum_{l=I+1-i}^{I-2} \text{Var}(C_{i,l+1} | D_l) \prod_{k=l+1}^{I-1} f_k^2 \\ &= \sum_{l=I+1-i}^{I-1} \sigma_l^2 E \left[\frac{C_{i,l}^2}{\delta_{i,l}} | D_l \right] \prod_{k=l+1}^{I-1} f_k^2 \\ &\quad + \text{Var}[C_{i,I+1-i} | D_I] \prod_{k=I+1-i}^{I-1} f_k^2 \\ &\quad + \sum_{l=I+2-i}^{I-1} \text{Var}[C_{i,l} | D_l] \prod_{k=l}^{I-1} f_k^2. \quad (\text{A.12}) \end{aligned}$$

Since $\text{Var}[C_{i,l+1}|D_l] = 0$ and from the fact that

$$\sum_{l=l+1-i}^{l-2} \text{Var}(C_{i,l+1}|D_l) \prod_{k=l+1}^{l-1} f_k^2 = \sum_{l=l+2-i}^{l-1} \text{Var}[C_{i,l}|D_l] \prod_{k=l}^{l-1} f_k^2,$$

we finally get the proof of Lemma A.1

A.5.2. Proof of Proposition A.1

Following Mack (1993), Mack (1994), we replace S_k^2 with $E(S_k^2|B_k)$ and $S_j S_k$ with $E(S_j S_k|B_k)$. This means that we approximate S_k^2 and $S_j S_k$ by varying and averaging as little data as possible so that as many values $C_{i,k}$ from data observed are kept fixed. Due to Proposition 4.1 (i) we have $E(\hat{f}_k - f_k) = 0$ and therefore $E(S_j S_k|B_k) = 0$ for $j < k$ because all f_r , $r < k$, are scalars under B_k . Since $E((\hat{f}_k - f_k)^2|B_k) = \text{Var}(\hat{f}_k|B_k)$ we obtain from (A.6)

$$E(S_k^2|B_k) = \hat{f}_{l+1-i}^2 \cdots \hat{f}_{k-i}^2 \text{Var}(\hat{f}_k|B_k) f_{k+1}^2 \cdots f_{l-1}^2.$$

Taken together, we have replaced $F^2 = \sum_{k=l+1-i}^{l-1} S_k^2$ with $\sum_{k=l+1-i}^{l-1} E(S_k^2|B_k)$ and the unknown parameters are replaced by their estimators. Altogether, we estimate F^2 by

$$\sum_{k=l+1-i}^{l-1} \hat{f}_{l+1-i}^2 \cdots \hat{f}_{k-1}^2 \hat{f}_k^2 \hat{f}_{k+1}^2 \cdots \hat{f}_{l-1}^2 \frac{\text{Var}(\hat{f}_k|B_k)}{\hat{f}_k^2}.$$

A.5.3. Proof of Proposition A.2

From definition of $\hat{f}_k$ in (14), we have

$$\begin{aligned} \text{Var}(\hat{f}_k|B_k) &= \text{Var}\left(\frac{\sum_{j=1}^{l-k} \gamma_{j,k} F_{j,k}}{\sum_{j=1}^{l-k} \gamma_{j,k}} \middle| B_k\right) \\ &= \frac{\sum_{j=1}^{l-k} \gamma_{j,k}^2 \text{Var}(F_{j,k}|B_k) \cdot \mathbf{1}_{\{\delta_{j,k} \neq 0\}}}{\left(\sum_{j=1}^{l-k} \gamma_{j,k}\right)^2}, \end{aligned}$$

where the second equality is due to the B_k -measurability of $\gamma_{j,k}$, the assumption (4.3) and the convention that the product of 0 and ∞ equals to 0.

A.5.4. Proof of Proposition A.3

We find the estimator $\widehat{F_i F_j}$ in the similar way to the estimator $\widehat{F^2}$ (see proof of Proposition A.1).

$$E[(S_k^i)^2|B_k] = \hat{f}_{l+1-i}^2 \cdots \hat{f}_{k-i}^2 \text{Var}(\hat{f}_k|B_k) f_{k+1}^2 \cdots f_{l-1}^2.$$

For $i < j$, we have

$$\begin{aligned} \widehat{F_i F_j} &= \sum_{k=l+1-i}^{l-1} \hat{f}_{l+1-j} \cdots \hat{f}_{l-i} \\ &\quad \cdot \hat{f}_{l+1-i}^2 \cdots \hat{f}_{k-i}^2 \text{Var}(\hat{f}_k|B_k) \hat{f}_{k+1}^2 \cdots \hat{f}_{l-1}^2 \\ &= \sum_{k=l+1-i}^{l-1} \frac{\text{Var}(\hat{f}_k|B_k)}{\hat{f}_k^2} \hat{f}_{l+1-j} \cdots \hat{f}_{l-i} \\ &\quad \cdot \hat{f}_{l+1-i}^2 \cdots \hat{f}_{k-1}^2 \cdot \hat{f}_k^2 \cdot \hat{f}_{k+1}^2 \cdots \hat{f}_{l-1}^2 \\ &= \sum_{k=l+1-i}^{l-1} \frac{\text{Var}(\hat{f}_k|B_k)}{\hat{f}_k^2} (\hat{f}_{l+1-j} \cdots \hat{f}_{l-1}) \\ &\quad \cdot (\hat{f}_{l+1-i} \cdots \hat{f}_{l-1}). \end{aligned} \tag{A.13}$$

B. Data

We present in Table B.8 the triangle of RAA data analysed in Mack (1994) and Murphy, Bardis, and Majidi (2012).

Table B.8. RAA run-off triangle (cumulative payments)

Accident Year i	Development Year j									
	1	2	3	4	5	6	7	8	9	10
1	5 012	8 269	10 907	11 805	13 539	16 181	18 009	18 608	18 662	18 834
2	106	4 285	5 396	10 666	13 782	15 599	15 496	16 169	16 704	
3	3 410	8 992	13 873	16 141	18 735	22 214	22 863	23 466		
4	5 655	11 555	15 766	21 266	23 425	26 083	27 067			
5	1 092	9 565	15 836	22 169	25 955	26 180				
6	1 513	6 445	11 702	12 935	15 852					
7	557	4 020	10 946	12 314						
8	1 351	6 947	13 112							
9	3 133	5 395								
10	2 063									

C. Individual link ratios of RAA run-off triangle

Table C.9. Individual link ratios F_{ij} (age-to-age factors) of run-off triangle RAA

AY	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10
1	1,650	1,319	1,082	1,147	1,195	1,113	1,033	1,003	1,009
2	40,425	1,259	1,977	1,292	1,132	0,993	1,043	1,033	
3	2,637	1,543	1,163	1,161	1,186	1,029	1,026		
4	2,043	1,364	1,349	1,102	1,113	1,038			
5	8,759	1,656	1,400	1,171	1,009				
6	4,260	1,816	1,105	1,226					
7	7,217	2,723	1,125						
8	5,142	1,887							
9	1,722								

D. LAD estimator

The least absolute deviation (LAD) method or L_1 (also known as Least Absolute Value (LAV)) method is a widely known alternative to the classical least squares (LS) or L_2 method for statistical analysis of linear regression models. Instead of minimizing the sum of squared errors, it minimizes the sum of absolute values of errors. More precisely, in the context of linear regression model, estimates are found by solving the following optimisation problem

$$\min_{\beta} \left\{ \sum_{i=1}^n |e_i| \right\} = \min_{\beta} \left\{ \sum_{i=1}^n \left| y_i - \sum_{j=1}^m x_{ij} \beta_j \right| \right\},$$

where $e_i := y_i - \sum_{j=1}^m x_{ij} \beta_j$, $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$. Unlike the LS method, the LAD method is not sensitive to outliers and produces robust estimates. LAD method is reduced to a linear programming problem and the computational difficulty is now entirely overcome by the availability of computing power and the effectiveness of linear programming.

Least absolute values (LAV) regression is very resistant to observations with unusual values in data.

In the numerical example presented in Section 6.2.2, we used one-dimensional ($m = 1$) LAD procedure where we took, for each column k of run-of triangle, $y_i := F_{k,i}$ and $x_{i,1} := 1$ for all i .

One more thing merits mentioning here. In the simple one-dimensional case ($m = 1$) the LAD estimator yields to the *sample median* (see Abur and Expósito 2004, p. 141).

E. Robust estimation—Limits in estimation of β parameters

We want to examine the existence of solution of equation (6.2). In this purpose, we study the properties of the flowing type of functions,

$$h_k(\beta_k) := \left(\sum_{i=1}^{N_k} a_{i,k}^{-\beta} b_{i,k} \right) \left(\sum_{i=1}^{N_k} a_{i,k}^{\beta} c_{i,k} \right),$$

with $a_{i,k} := C_{i,k}$, $b_{i,k} := \gamma_{i,k}^2 / (\sum_{i=1}^{I-k} \gamma_{i,k})^2$, $c_{i,k} := 1/(I - k - 1)$ ($\hat{f}_k - F_{i,k}$) and $N_k := I - k$. In the sequel, without the loss of generality, we omit the index k corresponding to the column of run-off triangle. Thus, we consider the function

$$h(\beta) := \left(\sum_{i=1}^N a_i^{-\beta} b_i \right) \left(\sum_{i=1}^N a_i^{\beta} c_i \right),$$

with $a_i \geq 0$, $0 \leq b_i \leq 1$ and $c_i \geq 0$. We rewrite the function h as follows:

$$h(\beta) := \sum_{i=1}^N \sum_{j=1}^N (a_i^{-\beta} b_i) (a_j^{\beta} c_j).$$

We easily observe that the function h tends to ∞ as β tends to ∞ or $-\infty$. Let us define, $d_{i,j} := (a_j/a_i)$, for $i < j$, and where the indices $i_k < j_k$ are such that $(a_j/a_i) > 1$. Then, by simple decomposition of double sum, we get:

$$h(\beta) := \sum_{i=1}^N b_i c_i + \sum_{i=1}^N \sum_{j=i+1}^N b_i c_j \cdot d_{i,j}^{\beta} + \sum_{i=1}^N \sum_{j=i+1}^N b_j c_i \cdot d_{j,i}^{\beta},$$

where by our notation $d_{i,j} > 1$ and $d_{j,i} < 1$. By the straightforward computations it is easy to show that

$$\begin{aligned} h'(\beta) &:= \sum_{i=1}^N \sum_{j=i+1}^N b_i c_j \cdot \ln(d_{i,j}) \cdot d_{i,j}^{\beta} \\ &\quad + \sum_{i=1}^N \sum_{j=i+1}^N b_j c_i \cdot \ln(d_{j,i}) \cdot d_{j,i}^{\beta}, \end{aligned}$$

Since $\ln(d_{i,j}) > 0$ and $\ln(d_{j,i}) < 0$, the first derivative h' has a limit in $-\infty$ and ∞ if β tends to $-\infty$ and ∞ respectively. In addition, the second derivative h'' is given by

$$\begin{aligned} h''(\beta) &:= \sum_{i=1}^N \sum_{j=i+1}^N b_i c_j \cdot (\ln(d_{i,j}))^2 \cdot d_{i,j}^{\beta} \\ &\quad + \sum_{i=1}^N \sum_{j=i+1}^N b_j c_i \cdot (\ln(d_{j,i}))^2 \cdot d_{j,i}^{\beta}. \end{aligned}$$

Given that $d_{i,j}^{\beta}$ and $d_{j,i}^{\beta}$ are strictly positive functions and all coefficients are positive, the second derivative of h is strictly positive. This means that first derivative of h is increasing function. Together with the previous facts it implies that h has an absolute minimum. In consequence, the equation (23) has zero, one or two solutions.

In the case where two solutions of opposite sign exist, the actuary should decide which one corresponds better to the considered line of business. In fact, as mentioned in Murphy, Bardis, and Majidi (2012), the choice of negative solution does not seem to be unreasonable in some situations. This issue is out of scope of this paper. In our numerical example the solution is determined by Excel tool called solver.

References

- Abur, A., and A. Expósito, *Power System State Estimation: Theory and Implementation*, Boca Raton, FL: CRC Press, 2004.
- Barnett, G., and B. Zehnwrith, "Best Estimates for Reserves," *Proceedings of the Casualty Actuarial Society* 87, Part 2, 2000, pp. 245–321.
- Blumsohn, G., and M. Laufer, "Unstable Loss Development Factors," *Casualty Actuarial Society E-Forum*, 2009, pp. 1–38.
- Buchwalder, M., H. Bühlmann, M. Merz, and M. V. Wüthrich, "The Mean Square Error of Prediction in the Chain Ladder

- Reserving Method (Mack and Murphy Revisited),” *ASTIN Bulletin* 36: 2, 2006a, pp. 521–542.
- Buchwalder, M., H. Bühlmann, M. Merz, and M. V. Wüthrich, “The Mean Square Error of Prediction in the Chain Ladder Reserving Method—Final Remark,” *ASTIN Bulletin* 36, 2006b, pp. 553–553.
- Huber, P. J., and E. M. Ronchetti, *Robust Statistics*, 2009.
- Jeng, H.-W., “On Small Samples and the Use of Robust Estimators in Loss Reserving,” Casualty Actuarial Society *E-Forum*, Fall 2010.
- Mack, T., “Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates,” *ASTIN Bulletin* 23, 1993, pp. 213–222.
- Mack, T., “Measuring the Variability of Chain Ladder Reserve Estimates,” Casualty Actuarial Society *Forum* 1, 1994, pp. 101–182.
- Mack, T., “The Standard Error of Chain Ladder Reserve Estimates Recursive Calculation and Inclusion of Tail Factor,” *ASTIN Bulletin* 29: 2, 1999, pp. 361–366.
- Mack, T., G. Quarg, and C. Braun, “The Mean Square Error of Prediction in the Chain Ladder Reserving Method: A Comment,” *ASTIN Bulletin* 36: 2, 2006, pp. 543–552.
- Merz, M., and M. V. Wüthrich, “Modelling the Claims Development Result for Solvency Purposes,” Conference Paper, *ASTIN Colloquium*, Manchester, July 2008.
- Murphy, D. M., “Unbiased Loss Development Factors,” *Insurance Mathematics and Economics* 18: 3, 1996, p. 228.
- Murphy, D., M. Bardis, and A. Majidi, “A Family of Chain-Ladder Factor Models for Selected Link Ratios,” *Variance* 6, 2012, pp. 143–160.
- Saito, S., “Generalisation of Mack’s Formula for Claims Reserving with Arbitrary Exponents for the Variance Assumption,” *Journal of Mathematics for Industry* 1, 2009, pp. 7–15.
- Sloma, P. *Contribution to the Weak Convergence of Empirical Copula Process. Contribution to the Stochastic Claims Reserving in General Insurance*. PhD thesis, Université Pierre et Marie Curie, 2014.
- Sloma, P., “General Model for Measuring the Uncertainty of the Claims Development Result (CDR),” *Proceedings of the Actuarial and Financial Mathematics Conference*, February 10–11, 2011.
- Wüthrich, M. V., M. Merz, and H. Bühlmann, “Bounds on the Estimation Error in the Chain Ladder Method,” *Scandinavian Actuarial Journal* 2008: 4, 2008, pp. 283–300.