Cash flow generalisations of non-life insurance expert systems estimating outstanding liabilities

Munir Hiabu
Munir.Hiabu.1@cass.city.ac.uk
Carolin Margraf
Carolin.Margraf.1@cass.city.ac.uk
María Dolores Martínez Miranda
mmiranda@ugr.es
Jens Perch Nielsen
Jens.Nielsen.1@city.ac.uk

Cass Business School, City University London, U.K.
September 17, 2015

Abstract

For as long as anyone remembers non-life insurance companies have used the so called chain ladder method to reserve for outstanding liabilities. When historical payments of claims are used as observations then chain ladder can be understood as estimating a multiplicative model. In most non-life insurance companies a mixture of paid data and expert knowledge, incurred data, is used as observations instead of just payments. This paper considers recent statistical cash flow models for asset-liability hedging, capital allocation and other management decision tools, and develops two new such methods incorporating available incurred data expert knowledge into the outstanding liability cash flow model. These two new methods unbundle the incurred data to aggregates of estimates of the future cash flow. By a re-distribution to the right algorithm, the estimated future cash flow is incorporated in the overall estimation process and considered as data. A statistical validation technique is developed for these two new methods and they are compared to the other recent cash flow methods. The two methods show to have a very good performance on the real-life data set considered.

Keywords: Stochastic Reserving; General Insurance; Chain Ladder; Claims Inflation; Incurred Data; Model Validation.

1 Introduction

The non-life insurance business is an important part of the economy for most developed countries with market revenues amounting to around five percent of GNP's. The best estimate of outstanding liabilities - often called the reserve - is perhaps the single most important number on the balance sheet of most non-life insurance companies. Insufficient reserves is one common reason for non-life insurance companies to go broke. A mismanaged reserving process can also lead to spurious and volatile yearly results leading to uninformed management decisions. Finally, a smoother and more transparent reserving process leads to significant cost savings in almost any non-life insurance company.

In this light it is perhaps surprising that statistical models for the most often used data set, the incurred run off triangle, are rarely considered in the literature. Incurred data is a mixture of historical payments of already settled claims and predicted severities of reported but not settled claims. The predicted severities are based on all available expert opinion in the company and are called case estimates. In that spirit, incurred data has added information about reported claims which are not available when only historical payments are considered. Actuaries often prefer working with chain ladder on the incurred triangle rather than on historical payments. But it is a bit unclear what this really means in terms of mathematical statistics. While incurred chain ladder probably makes good sense in a deterministic forecasting framework, the stochastic nature of the expert opinion in incurred data is not taken into account by this practice.

There is a little literature acknowledging the added value of incurred data: the probably most famous Munich chain ladder approach by Quarg and Mack (2004), regression approaches by Halliwell (1997), Halliwell (2009), Venter (2008), and a paid-incurred chain reserving method by Posthuma et al. (2008), Merz and Wüthrich (2010), Happ et al. (2012) and Happ and Wüthrich (2013). These eight papers combine payment data and incurred data into one statistical model resulting in one reserve estimate. However, none of those papers take the common micro-structure of payment data and incurred data into account. Payment data and incurred data are both constructed via manipulations on the same underlying data; incurred data with the extra element of expert knowledge on expected future cash flows.

To model this relationship one needs micro-structural assumptions about the underlying claim process. Pigeon et al. (2014) does exactly this. It is based on a recent trend to use so called granular data or micro data for reserving, see Antonio and Plat (2014) for one of the most interesting recent contributions in that area. See also Martínez-Miranda et al. (2013a) for a continuous interpretation of the classical chain ladder methodology. While these approaches indeed seem to be favorable, they are not well established yet and rely on data and granular information that is most often not available at hand.

Double chain ladder, introduced in Martínez-Miranda et al. (2012), builds on micro-structural assumptions but does not need granular data in the estimation procedure. It is based on the methods of Verrall et al. (2010) and Martínez-Miranda et al. (2011) where the objective was to only rely on data that is already available in most reserving departments. It uses additional information of claim counts which is another triangle most often available in the data portfolio of a reserving department. The result is a full statistical model based on historical payments and counts capable of incorporating the information of the incurred data in a natural way.

Agbeko et al. (2015) recently introduced a model reproducing the deterministic results of the earlier mentioned chain ladder method on incurred data by incorporating the incurred data expert opinion into the well-defined full stochastic model of double chain ladder. One direct advantage of this approach is that the chain ladder model based on paid data can be validated against the chain ladder model based on incurred data. While the paid chain ladder and incurred chain ladder methods have been available for a long time as part of almost any non-life actuary's tool kit, it has never before been possible to compare them against each other when only the typical aggregated data were available.

Two other methods of incorporating expert knowledge of incurred data into these full cash flow models have been introduced in Martínez-Miranda et al. (2013b) and Hiabu et al. (2016). The first of these two methods is extracting the inflation of the cost of a single claim from the incurred data and then incorporates that information in the double chain ladder model of Martínez-Miranda et al. (2012). The second of these two methods suggests to incorporate a RBNS-preserving property. RBNS stands for Reported But Not yet Settled, and it can be estimated by the sum of all case estimates. This estimate is the best the claims department of an insurance company (with all the expert knowledge on the nature and severity of each claim available in such a department) is able to do. Hiabu et al. (2016) therefore produced a version of double chain ladder reproducing exactly the expert judgement of the RBNS reserves.

All those mentioned double chain ladder extensions take advantage of the underlying structure of the incurred data, extract the relevant information from it and plug it into the original double chain ladder method. The advantage of this approach is that the simplicity and intuition of the simple chain ladder method is preserved and that the full statistical interpretation and stochastic cash flow formulation is inherited from double chain ladder.

The two new stochastic cash flow methods developed in this paper build on the ideas and techniques of Hiabu et al. (2016). The first one treats the expert knowledge of the incurred data as real data and incorporates it into the model, the second builds a second RBNS preserving cash flow model on top of this method. The approach is first to unbundle the incurred data to get back to the original aggregated RBNS numbers. These aggregated numbers are re-distributed according to the estimated delay such that the resulting algorithm takes both historical data and expert data into account in the final estimation. We therefore let the estimated future cash flow be incorporated in the overall estimation process by considering it as data. Note that both of these two new methods are cash flow models of the same nature as the models considered in Agbeko et al. (2015), and they can therefore be validated and compared to the models considered there. In the applied data example, this validation indicates, that the two new methods seem to take better advantage of the incurred expert data than previous methods did.

Recent years have seen a growing interest in expert systems related to non-life insurance, see for example Belles-Sampera et al. (2014) and Abbasi and Guillén (2013), who consider ways of understanding risk in non-life insurance. Guelman and Guillén (2014) work with pricing of insurance claims and the customers sensitivity to that price; Guillen et al. (2012) and Kaishev et al. (2013) transfer knowledge from one business line to another to optimize cross-selling. Human judgement is important in all these insurance applications. When prices are set, there is a business intelligence department evaluating how much weight to put on the model at hand and how much weight to put on market prices as such. When risk is evaluated, human judgement calibrates the entering parameters. And when RBNS claims reserves are set, then there is an element of human judgement in the settlement of every single claim. It is also a human judgement when it is decided to use model based claims reserves for some subset of the claims, for example the smaller ones.

We conclude this introduction by noting that Martínez-Miranda et al. (2012) has two versions of double chain ladder; one version where the delay is not adjusted and another where it is adjusted. In this paper only the unadjusted version of double chain ladder is considered. One reason for the adjustment of the delay in Martínez-Miranda et al. (2012) was to improve the performance of estimating the out-of-sample tail reserve. While this is a very important issue, it is beyond the scope of this paper to consider the out-of-sample tail reserve.

The rest of the paper is structured as follows. Section 2 describes the data and the expert knowledge, introduces the notation and defines the model assumptions. Section 3 discusses the outstanding loss liabilities point estimates. Section 4 describes four methods to estimate the parameters in the model: DCL, BDCL, PDCL, IDCL, EDCL and PEDCL. An application is considered in Section 5 and the validation of the six methods against each other is gone through in Section 6. Finally, Section 7 provides some concluding remarks.

2 Data and first moment assumptions

This chapter introduces the data used in maybe every insurance reserving department to calculate their outstanding liabilities. Also the methods described in this paper rely on these data sets. They are often shortly called run-off triangles. These run-off triangles are the aggregated incurred counts (data), aggregated payments (data) and aggregated incurred payments (mixture of data and expert knowledge). All of those three objects have the same structural form, i.e., they live on the upper triangle

$$\mathcal{I} = \{(i,j) : i = 1,\ldots,m, j = 0,\ldots,m-1;\ i+j \leq m\}, \quad m > 0,$$

where m is the number of underwriting years observed. The parameter m also has another crucial role. If no tail-factors are considered, which will be assumed throughout this paper, then $m-1$ is the maximum delay, that is the time from the underwriting of the policy a claim is based on until its payment. This assumption is called being run-off, hence the name run-off triangles.
Let's first introduce the two data triangles.

Aggregated incremental incurred counts: $N_{\mathcal{I}} = \{N_{ik} : (i,k) \in \mathcal{I}\}$, with N_{ik} being the total number of claims of insurance incurred in year i which have been reported in year $i+k$, i.e. with k periods delay from year i.

Aggregated incremental payments: $X_{\mathcal{I}} = \{X_{ij} : (i,j) \in \mathcal{I}\}$, with X_{ij} being the total payments from claims incurred in year i and paid with j periods delay from year i.

A often confusing point is that the meaning of the second coordinate of the triangle $\mathcal{I}$ varies between the two different data. While in the counts triangle it represents the reporting delay, in the payments triangle it represents the development delay, that is reporting delay plus settlement delay.
The definition of the incurred payments triangles is not that straight forward. To allow for a exact description, we first introduce micro structural variables of the claims process. We hereby follow the line of Martínez-Miranda et al. (2012), since those variables will also play a key role in the underlying assumption of the double chain ladder model.
By N_{ikl}^{paid}, we denote the number of the future payments originating from the N_{ik} reported claims, which were finally paid with a delay of $k+l$, where $l = 0,\ldots,m-1$. Also, let $X_{ikl}^{(h)}$ denote the individual settled payments which arise from N_{ikl}^{paid}, $h = 1,\ldots,N_{ikl}^{paid}$. Finally, we define

$$X_{ikl} = \sum_{h=0}^{N_{ikl}^{paid}} X_{ikl}^{(h)}, \quad (i,k) \in \mathcal{I},\ l = 0,\ldots,m-1,$$

i.e., those payments originating from underwriting year i, which are reported after a delay of k and paid with an overall delay of $k+l$.
The aggregated incurred payments are then considered as unbiased estimators of $\sum_{l=0}^{m-1} X_{ikl}$. Technically, we model the expert knowledge as follows.

Expert knowledge:

Aggregated incurred payments: $I_{\mathcal{I}} = \{I_{ik} : (i,k) \in \mathcal{I}\}$, with I_{ik} being

$$I_{ik} = \sum_{s=0}^{k} \sum_{l=0}^{m-1} E[X_{isl}|\ \mathcal{F}_{(i+k)}] - \sum_{s=0}^{k-1} \sum_{l=0}^{m-1} E[X_{isl}|\ \mathcal{F}_{(i+k-1)}],$$

where $\mathcal{F}_h$ is an increasing filtration illustrating the expert knowledge at calendar period h.

In this manuscript, we will only consider best estimates and therefore only need assumptions on the mean. We show that the classical CLM multiplicative structure holds under very general underlying dependencies on the mean. The first moment conditions of the DCL model are formulated below. For fixed $i = 0, \ldots, m;\quad k, l = 0, \ldots, m-1,$ and $h = 1, \ldots, N_{ikl}^{paid}$, it holds that

A1. The counts N_{ik} are random variables with mean having a multiplicative parametrization $\mathrm{E}[N_{ik}] = \alpha_i \beta_k$, and identification $\sum_{k=0}^{m-1} \beta_k = 1$.

A2. The mean of the RBNS delay variables is $\mathrm{E}[N_{ikl}^{paid} | N_{\mathcal{I}}] = N_{ik} \widetilde{\pi}_l$.

A3. The mean of the individual payments size conditional on the number of payments and the counts is given by $\mathrm{E}[X_{ikl}^{(h)} | N_{ikl}^{paid}, N_{\mathcal{I}}] = \widetilde{\mu}_l \gamma_i$.

Assumption A1 is the classical chain ladder assumption applied on the counts triangle, see also Mack (1991). The main point hereby is the multiplicativity between underwriting year and reporting delay. Assumptions A2 and A3 are necessary to connect reporting delay, settlement delay and development delay - the main idea of DCL. See also Verrall et al. (2010), Martínez-Miranda et al. (2011) and Martínez-Miranda et al. (2012).

Using A1 to A3, we have that

$$\begin{aligned}
\mathrm{E}\left[\sum_{h=1}^{N_{i,j-l}^{paid}} X_{i,j-l,l}^{(h)} | N_{\mathcal{I}} \right] &= \mathrm{E}\left[\sum_{h=1}^{N_{i,j-l,l}^{paid}} \mathrm{E}[X_{i,j-l,l}^{(h)} | N_{\mathcal{I}}, N_{i,j-l,l}^{paid}] | N_{\mathcal{I}} \right] \\
&= \mathrm{E}[N_{i,j-l,l}^{paid} \widetilde{\mu}_{j-l} \gamma_i | N_{\mathcal{I}}] = N_{i,j-l} \widetilde{\pi}_l \widetilde{\mu}_{j-l} \gamma_i.
\end{aligned}$$

Note that the observed aggregated payments can be written as

$$X_{ij} = \sum_{l=0}^{j} X_{i,j-l,l} = \sum_{l=0}^{j} \sum_{h=1}^{N_{i,j-l,l}^{paid}} X_{i,j-l,l}^{(h)}.$$

With the previous consideration, we derive

$$\mathrm{E}[X_{ij}|N_{\mathcal{I}}] = \gamma_i \sum_{l=0}^{k} N_{i,j-l} \widetilde{\pi}_l \widetilde{\mu}_{j-l,l},$$

and the unconditional mean is

$$\mathrm{E}[X_{ij}] = \alpha_i \gamma_i \sum_{l=0}^{j} \beta_{j-l} \widetilde{\mu}_{j-l,l} \widetilde{\pi}_l. \tag{1}$$

Inspecting equation (1), we can reduce the amount of parameters by setting $\mu = \sum_{l=0}^{m-1} \widetilde{\pi}_l \widetilde{\mu}_l$ and $\pi_l = \widetilde{\pi}_l \widetilde{\mu}_l \mu^{-1}$, so that $\mu \pi_l = \widetilde{\mu}_l \widetilde{\pi}_l$ and therefore the unconditional mean of the payments becomes

$$\mathrm{E}[X_{ij}] = \alpha_i \gamma_i \mu \sum_{l=0}^{j} \beta_{j-l} \pi_l. \tag{2}$$

Equation (2) is the key in deriving the outstanding loss liabilities. These are the values of (X_{ij}) in the lower triangle. Consequently in the sequel,

$$(\alpha, \beta, \pi, \gamma, \mu) = (\alpha_1, \ldots, \alpha_m, \beta_0, \ldots, \beta_{m-1}, \pi_0, \ldots, \pi_{m-1}, \gamma_1, \ldots, \gamma_m, \mu)$$

are called the DCL parameters. In the next section, we will see that in a very natural way, we are able to distinguish between RBNS and IBNR claims. This is possible due to the separation of the development delay into the reporting delay β and the settlement delay π.

3 Forecast outstanding claims: the RBNS and IBNR reserves and predictive distributions

In this section, we assume that the DCL parameters $(\alpha, \beta, \pi, \gamma, \mu)$ are already derived and show how easily point forecasts of the RBNS and IBNR components of the reserve can be calculated. Note that when calculating the RBNS part, it is possible to replace the parameter (α_i, β_k) by the true value N_{ik}, since the claims are already reported and thus N_{ik} is observed. However, for the IBNR reserves, it is obviously necessary to use all DCL parameters, including the estimates of future numbers of incurred claims $\alpha_i \beta_k$.

Using the notation of Verrall et al. (2010) and Martínez-Miranda et al. (2011), we consider predictions over the triangle, $\mathcal{J} = \mathcal{J} = \{i = 2, \ldots, m; j = 0, \ldots, m - 1 \text{ with } i + j \geq m + 1\}$, illustrated in Figure 1.

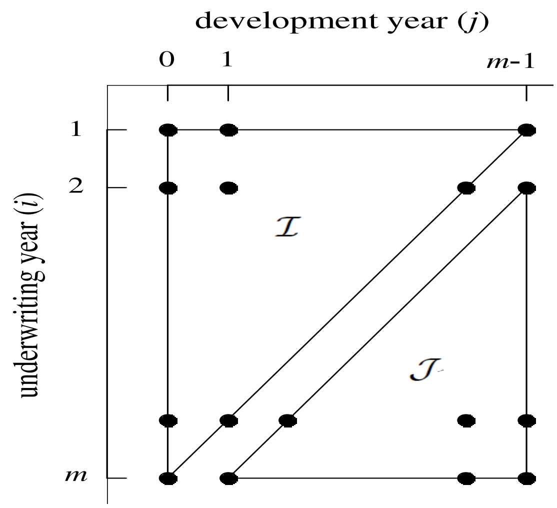

Figure 1: Index sets for aggregate claims data, assuming a maximum delay of $m-1$.

We define the RBNS component as follows, where we consider two possibilities depending on whether the estimates of N_{ik} are used or not.

$$\widehat{X}_{ij}^{rbns(1)} = \sum_{l=i-m+j}^{j} N_{i,j-l} \widehat{\pi}_l \widehat{\mu} \widehat{\gamma}_i, \quad (i, j) \in \mathcal{J}. \tag{3}$$

and

$$\widehat{X}_{ij}^{rbns(2)} = \sum_{l=i-m+j}^{j} \widehat{N}_{i,j-l} \widehat{\pi}_l \widehat{\mu} \widehat{\gamma}_i, \quad (i, j) \in \mathcal{J}, \tag{4}$$

where $\widehat{N}_{ij} = \widehat{\alpha}_i \widehat{\beta}_j$. In most cases, to shorten the notation, we will simply write $\widehat{X}_{ij}^{rbns}$ for the RBNS estimates. However, whenever it is necessary, we will state which version is taken. The IBNR component always needs all DCL parameters:

$$\widehat{X}_{ij}^{ibnr} = \sum_{l=0}^{i-m+j-1} \widehat{N}_{i,j-l} \widehat{\pi}_l \widehat{\mu} \widehat{\gamma}_i, \quad (i, j) \in \mathcal{J}. \tag{5}$$

The outstanding loss liabilities point estimates are then,

$$\widehat{X}_{ij} = \widehat{X}_{ij}^{rbns} + \widehat{X}_{ij}^{ibnr} \tag{6}$$

The outstanding liabilities per accident year are the row sums of IBNR and RBNS estimates. For a fixed i, we write $\mathcal{J}(i) = \{j : (i, j) \in \mathcal{J}\}$. Then the outstanding liabilities per accident year $i = 1, \ldots, m$ are

$$\widehat{R}_i = \sum_{j \in \mathcal{J}(i)} \widehat{X}_{ij}^{rbns} + \widehat{X}_{ij}^{ibnr}. \tag{7}$$

In the next section we describe several methods to derive the DCL parameters.

4 Estimation of the parameters in the Double Chain Ladder model

To estimate the outstanding claims and thereby construct RBNS and IBNR reserves, we need to estimate the parameters involved in (2). In this section, we explore six different estimators.

4.1 The DCL method

The DCL method is the original and maybe most simple method to derive the parameters introduced in the previous section. It was introduced in Martínez-Miranda et al. (2012). To estimate the DCL parameters in (2), assumptions on the payments triangle $X_{\mathcal{I}}$ are needed. DCL assumes the assumptions underlying the CLM method.

B1 The payments X_{ij}, with $i = 1, \ldots, m$, and $j = 0, \ldots, m-1$, are random variables with mean having a multiplicative parametrization:

$$E[X_{ij}] = \widetilde{\alpha}_i \widetilde{\beta}_j, \qquad \sum_{j=0}^{m-1} \widetilde{\beta}_j = 1. \tag{8}$$

Finally, merging (2) and (8), we conclude

$$\alpha_i \gamma_i \mu \sum_{l=0}^{j} \beta_{j-l} \pi_l = \widetilde{\alpha}_i \widetilde{\beta}_k,$$

and identify the parameters by

$$\alpha_i \mu \gamma_i = \widetilde{\alpha}_i, \tag{9}$$

$$\sum_{l=0}^{j} \beta_{j-l} \pi_l = \widetilde{\beta}_j. \tag{10}$$

Again, many other micro-structure formulations might exist, thus the one specified by (9) and (10) is only one of several possible. However, the above model can be considered as a detailed specification of the CLM. In Martínez-Miranda et al.

(2013b) it is shown that if the RBNS component is calculated by (4), DCL completely replicates the results of CLM.

Now, the main idea to derive the DCL parameters is to estimate the chain ladder parameters $(\widehat{\alpha}, \widehat{\beta})$ and $(\widehat{\widetilde{\alpha}}, \widehat{\widetilde{\beta}})$ (cf. A1, B1) by applying the classical chain ladder algorithm on the payments triangle $X_{\mathcal{I}}$ and the counts triangle $N_{\mathcal{I}}$. Afterwards, the parameters left in (2) (this is $(\widehat{\gamma}, \widehat{\mu}, \widehat{\pi})$) can be calculated by simple algebra using (9) and (10).

For illustration of the chain ladder algorithm, we assume an incremental triangle (C_{ij}) (in our case this would be $N_{\mathcal{I}}$ or $X_{\mathcal{I}}$), and that we want to estimate its chain ladder parameters $(\widehat{\overline{\alpha}}, \widehat{\overline{\beta}})$. To apply the chain ladder algorithm, one has to transform the triangle (C_{ij}) into a cumulative triangle (D_{ij}):

$$D_{ij} = \sum_{k=1}^{j} C_{ik}.$$

Then, the chain ladder algorithm can be applied on (D_{ij}). It will produce estimates of development factors, λ_j, $j = 1, 2, \ldots, m-1$, which can be described by

$$\widehat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} D_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}}.$$

These development factors can be converted into estimates of $(\overline{\alpha}, \overline{\beta})$ using the following identities which were derived in Verrall (1991).

$$\widehat{\overline{\beta}}_0 = \frac{1}{\prod_{l=1}^{m-1} \widehat{\lambda}_l}$$

$$\widehat{\overline{\beta}}_j = \frac{\widehat{\lambda}_j - 1}{\prod_{l=j}^{m-1} \widehat{\lambda}_l}$$

$$\widehat{\overline{\alpha}}_i = \sum_{j=0}^{m-i} C_{ij} \prod_{j=m-i+1}^{m-1} \widehat{\lambda}_j$$

Alternatively, analytical expressions for the estimators can also be derived directly (rather than using the chain ladder algorithm) and further details can be found in Kuang et al. (2009).

Once the chain ladder parameters $(\widehat{\alpha}, \widehat{\beta})$ and $(\widehat{\widetilde{\alpha}}, \widehat{\widetilde{\beta}})$ are derived, the settlement delay parameter π can be estimated just by solving the following linear system.

$$\begin{pmatrix} \widehat{\widetilde{\beta}}_0 \\ \vdots \\ \vdots \\ \widehat{\widetilde{\beta}}_{m-1} \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_0 & 0 & \cdots & 0 \\ \widehat{\beta}_1 & \widehat{\beta}_0 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \widehat{\beta}_{m-1} & \cdots & \widehat{\beta}_1 & \widehat{\beta}_0 \end{pmatrix} \begin{pmatrix} \pi_0 \\ \vdots \\ \vdots \\ \pi_{m-1} \end{pmatrix}. \tag{11}$$

Let $\widehat{\pi}$ denote the solution of (11).

Now we consider the estimation of the parameters involved in the means of individual payments. Of course, the model is technically over-parametrised since there are too many inflation parameters in (9). The simplest way to ensure identifiability is to set $\gamma_1 = 1$, and then the estimate of μ, $\widehat{\mu}$, can be obtained from

$$\widehat{\mu} = \frac{\widehat{\widetilde{\alpha}}_1}{\widehat{\alpha}_1}. \tag{12}$$

Using $\widehat{\mu}$, the remaining estimates for $\gamma_i,\ i = 2, \dots, m$, are directly derived from (9). The estimation procedure of double chain ladder is already programmed with the language R. We have used the R-package DCL (Martínez-Miranda et al., 2013c) to derive Table 1, which shows the values of $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\pi}$ and $\widehat{\gamma}$ calculated from real data included in the DCL package.

i	k,l	$\widehat{\alpha}(i)$	$\widehat{\beta}(k)$	$\widehat{\pi}(l)$	$\widehat{\gamma}(i)$
1	0	1078	0.7599	0.0592	1.0000
2	1	1890	0.2097	0.3098	1.1173
3	2	2066	0.0189	0.2032	1.4947
4	3	2353	0.0064	0.1996	1.7461
5	4	3015	0.0016	0.1388	2.1075
6	5	3727	0.0010	0.0440	2.0936
7	6	5057	0.0009	0.0227	2.2495
8	7	6483	0.0007	0.0095	2.1250
9	8	7727	0.0003	0.0018	1.9028
10	9	7134	0.0001	0.0029	2.0197
11	10	7319	0.0001	0.0002	2.0704
12	11	6152	0.0000	0.0026	2.2666
13	12	5242	0.0001	0.0019	2.3157
14	13	6150	0.0000	0.0032	2.4747
15	14	7028	0.0001	-0.0002	2.3829
16	15	6725	0.0000	0.0013	2.8391
17	16	5260	0.0000	-0.0004	3.1815
18	17	5869	0.0000	0.0000	4.1747
19	18	5953	0.0000	0.0000	6.7501
			$\widehat{\mu} = 2579$		

Table 1: DCL parameter estimates derived by the DCL method

4.2 The BDCL method

The CLM and Bornhuetter-Ferguson (BF) methods are among the easiest claim reserving methods and, due to their simplicity, they are two of the most commonly used techniques in practice. Some recent papers on the BF method include Verrall (2004), Mack (2008), Schmidt and Zocher (2008), Alai et al. (2009) and Alai et al. (2010). The BF method introduced by Bornhuetter and Ferguson (1972) aims to address one of the well known weaknesses of CLM, which is the effect outliers can have on the estimates of outstanding claims. Especially the most recent underwriting years are the years with nearly no data and thus very sensitive to outliers. However, these recent underwriting years build the very major part of the outstanding claims. Hence, the CLM estimates of the outstanding liabilities might differ fatally from the true (unknown) values.
Acknowledging this problem, the BF method incorporates prior knowledge from experts and is therefore more robust than the CLM method, which relies completely on the data contained in the run-off triangle $X_{\mathcal{I}}$.

In this section, we briefly summarize the Bornhutter-Ferguson double chain ladder (BDCL) method introduced in Martínez-Miranda et al. (2013b), which mimics BF in the framework of DCL. The BDCL method starts with identical steps as DCL but instead of using the estimate of the inflation parameters, γ and μ, from the triangle of paid claims $X_{\mathcal{I}}$, it deploys expert knowledge in the form of the incurred triangle $I_{\mathcal{I}}$ to adjust the estimation of the sensitive inflation parameter γ. This is done as follows. From assumptions $A2, A3$ and equation (9), we easily deduce that

$$E[I_{ik}] = \alpha_i \mu \gamma_i \beta_k = \widetilde{\alpha}_i \beta_k. \tag{13}$$

Hence, the incurred triangle $I_{\mathcal{I}}$ has multiplicative mean and its underwriting year factor, $\widetilde{\alpha}$, is identical to the one of the payments triangle $X_{\mathcal{I}}$ (cf. (8)). However, its estimation is less sensitive to outliers since it incorporates all incurred claims via expert knowledge. We conclude that we can replace the payments triangle by the incurred payments triangle when we calculate estimates of the inflation parameters, γ, μ, in (9). Note that the severity mean μ is going to remain the same since the first rows of $X_{\mathcal{I}}$ and $I_{\mathcal{I}}$ are identical.

Summarized, the BDCL-method can be carried out as follows.

- *Step 1: Parameter estimation.*
  Estimate the DCL parameters $(\alpha, \beta, \pi, \gamma, \mu)$ using the DCL method of Section 4.1 with the data in the triangles $N_{\mathcal{I}}$ and $X_{\mathcal{I}}$ and denote the parameter estimates by $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}, \widehat{\gamma}, \widehat{\mu})$.

Repeat this estimation using the DCL method but replacing the triangle of paid claims $X_{\mathcal{I}}$ by the triangle of incurred data $I_{\mathcal{I}}$. Keep only the resulting estimated inflation parameters, denoted by $\widehat{\gamma}^{BDCL}$.

- *Step 2: BF adjustment.*
  Replace the inflation parameters $\widehat{\gamma}$ from the paid data by the estimate from the incurred triangle, $\widehat{\gamma}^{BDCL}$.

From Step 1 and Step 2, the final BDCL estimates of the DCL parameters are $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}, \widehat{\gamma}^{BDCL}, \widehat{\mu})$.

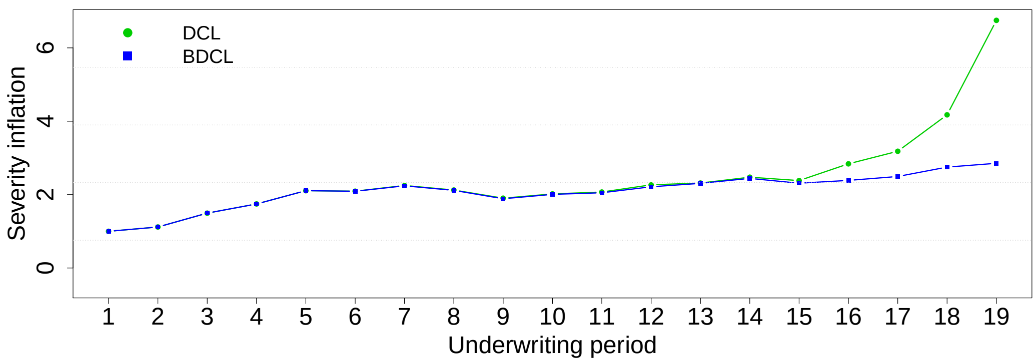

Figure 2: Plot of severity inflation estimates. DCL: $\widehat{\gamma}_i$ (green), BDCL: $\widehat{\gamma}_i^{BDCL}$ (blue).

Using the R-package DCL (Martínez-Miranda et al., 2013c), we derive Figure 2 which shows the severity inflation estimates derived by DCL and BDCL. BDCL, with the incorporated expert knowledge, seems to stabilize the severity inflation in the most recent underwriting years while keeping the values in the other years. The result is a more realistic estimate correcting the DCL parameter $\widehat{\gamma}_i$ exactly in its weakest point, that is in those years where the payments triangle $X_{\mathcal{I}}$ has nearly no data. Again, those recent underwriting years contain the very major part of the outstanding liabilities.

4.3 The IDCL method

This section gives a brief theoretical introduction to the IDCL method of Agbeko et al. (2015). In the BDCL definition, we incorporated an additional triangle of incurred claims in order to produce a more stable estimate of the underwriting inflation parameter γ_i. The derived BDCL method becomes a variant of the Bornhuetter-Ferguson technique using prior knowledge contained in the incurred triangle. One natural question is whether the derived reserve estimate is the classical incurred chain ladder. The answer is that this not the case; the IDCL method does not replicate the results obtained by applying the classical chain ladder method to the incurred triangle. Among practitioners, the incurred reserve seems to be more realistic for many datasets compared to the classical paid chain ladder reserve. IDCL mimics the reserve estimate of chain ladder on the incurred triangles in the DCL framework. It is defined just by rescaling the underwriting inflation parameter estimate from the DCL method. Specifically, we define a new scaled inflation factor estimate $\widehat{\gamma}^{IDCL}$ such that

$$\widehat{\gamma}_i^{IDCL} = \frac{\widehat{R}_i^*}{\widehat{R}_i}\widehat{\gamma}_i, \tag{14}$$

where R_i^* is the outstanding loss liabilities per accident year as predicted by applying the traditional CLM on incurred data, and $(R_i,\widehat{\gamma}_i)$ are the outstanding loss liabilities per accident year and the inflation parameter respectively, using the DCL method (cf. Section 4.1).

The final IDCL estimates of the DCL parameters are then $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}, \widehat{\gamma}^{IDCL}, \widehat{\mu})$. With the new inflation parameter estimate $\widehat{\gamma}^{IDCL}$, the outstanding liabilities derived by the DCL parameters completely replicate the CLM estimates on the incurred triangle.

4.4 The PDCL method

This section gives a brief theoretical introduction to the PDCL method introduced in Hiabu et al. (2016). See also Nielsen (2016). In the last section, we have described a method which incorporates expert knowledge in form of the incurred triangle $I_{\mathcal{I}}$. The values in $I_{\mathcal{I}}$ arise from case estimates for RBNS claims, developed in the case department of the insurance company, and claims which are already paid. Thus, if one subtracts these already paid claims (which are given via the payments triangle $X_{\mathcal{I}}$) from the incurred triangle, one can reconstruct the RBNS case estimates. However, as soon as this is done, it is obvious that these RBNS case estimates do not match with the RBNS estimates (3) and (4), using any DCL method (including BDCL). We conclude that the reserve department, using double chain ladder (and also chain ladder), calculates different RBNS estimates than those given by the case department.

PDCL replaces the calculated RBNS estimates with the case estimates. The method, however, does not only preserve the RBNS case estimates by adding the IBNR estimates to conclude the reserve, it also takes the RBNS case estimates to correct the other DCL parameters. Therefore, also the total IBNR size will change.

The first step is to construct a preliminary square (S_{ij}), $i = 1, \ldots, m$, $j =$

$0, \ldots, m-1$, which yields new estimators for the DCL parameters. The upper triangle of the square (i.e., $(i,j) \in \mathcal{I}$) should have the same entries as the payments triangle (X_{ij}). The lower triangle (i.e., $(i,j) \in \mathcal{J}$) should consist of preliminary estimates of the outstanding loss liabilities. The outstanding loss liabilites comprise an RBNS and an IBNR part (cf. (6)). However, only the IBNR part of these outstanding loss liabilities is estimated. For the RBNS component, the RBNS case estimates are taken. More precisely, one takes the DCL parameter estimates $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}, \widehat{\gamma}, \widehat{\mu})$ and use these parameters to estimate the RBNS component $(\widehat{X}_{ij}^{rbns})$ and IBNR component $(\widehat{X}_{ij}^{ibnr})$ using (4) and (5). As mentioned above, the RBNS estimate should be be equal to the RBNS case estimates, which can only be reconstructed per accident year. For $i = 1, \ldots, m$, they can be described as

$$X_i^{rbns.case.estimate} = \sum_{j=0}^{m-i} I_{ij} - \sum_{j=0}^{m-i} X_{ij}. \tag{15}$$

Hence, we define the RBNS preserving components

$$\widehat{X}_{ij}^{rbns.pres} = \frac{\sum_{j=0}^{m-i} I_{ij} - \sum_{j=0}^{m-i} X_{ij}}{\sum_{j \in \mathcal{J}(i)} \widehat{X}_{ij}^{rbns}} \widehat{X}_{ij}^{rbns}. \tag{16}$$

Note that

$$\sum_{j \in \mathcal{J}(i)} \widehat{X}_{ij}^{rbns.pres} = X_i^{rbns.case.estimate}.$$

Thus we define the preliminary square (S_{ij}) as

$$S_{ij} = \begin{cases} X_{ij}, & \text{if } (i,j) \in \mathcal{I}, \\ \widehat{X}_{ij}^{rbns.pres} + \widehat{X}_{ij}^{ibnr}, & \text{if } (i,j) \in \mathcal{J}. \end{cases} \tag{17}$$

One can easily see that payments square (S_{ij}) has approximately multiplicative mean $E[S_{ij}] \approx \widetilde{\alpha}_i \widetilde{\beta}_j$. Therefore, we can use (S_{ij}) to completely replace $X_{\mathcal{I}}$ to estimate the DCL parameters (cf. (8)). Note that in the BDCL method we were only able to balance the estimator of the inflation parameter $\widetilde{\alpha}_i$ (cf. (13)). Again, while in the BDCL method, one uses the expert knowledge to only adjust the inflation parameters, here, one can take full advantage of the triangle $I_{\mathcal{I}}$ and also correct the delay parameters.

Since (S_{ij}) has a multiplicative structure, the CLM idea is used to estimate $\widetilde{\alpha}_i$ and $\widetilde{\beta}_j$. We define

$$\widehat{\widetilde{\alpha}}_i^{PDCL} = \sum_{j=0}^{m-1} S_{ij}, \qquad \widehat{\widetilde{\beta}}_j^{PDCL} = \frac{\sum_{i=1}^{m} S_{ij}}{\sum_{(i,j) \in \mathcal{I} \cup \mathcal{J}} S_{ij}}. \tag{18}$$

Exactly as in the previous sections, one can now apply (9) and (10) to derive the PDCL parameters $(\widehat{\alpha}_i, \widehat{\beta}_j, \widehat{\pi}^{PDCL}, \widehat{\gamma}^{PDCL^*}, \widehat{\mu}^{PDCL})$. Since this approach is still not RBNS preserving, $\widehat{\gamma}^{PDCL^*}$ is balanced by defining a new scaled inflation factor estimate $\widehat{\gamma}^{PDCL}$ such that

$$\widehat{\gamma}^{PDCL} = \frac{\sum_{j=0}^{m-i} I_{ij} - \sum_{j=0}^{m-i} X_{ij}}{\widehat{X}_{ij}^{rbns}}, \tag{19}$$

where $\widehat{X}_{ij}^{rbns}$ is calculated with the parameters $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}^{PDCL}, \widehat{\gamma}^{PDCL^*}, \widehat{\mu}^{PDCL})$ using (4).

4.5 The EDCL method

In this chapter we introduce a new method called expert double chain ladder (EDCL), indicating that expert knowledge in form of incurred data and RBNS's are incorporated into the system as pseudo data. The idea of the EDCL method is to replicate the basic steps of the previously introduced PDCL method (15)-(18), but without the adjustment of the severity inflation in (19). Instead those steps are iterated until convergence. The iteration forces a homogeneous solution which incorporates the incurred triangle, I, and the payment triangle, X, to one reserve. The discrepancy between estimated RBNS and RBNS provided by the case estimates can be explained by variation of the observations around their mean. Note that given the model assumptions, we are estimating the mean of the RBNS.

The EDCL estimation can be described as follows. In the first step of the iteration, we start with the DCL parameter estimates $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}, \widehat{\gamma}, \widehat{\mu})$, which we denote by $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}^{EDCL,(0)}, \widehat{\gamma}^{EDCL,(0)}, \widehat{\mu}^{EDCL,(0)})$.

In the $k-$th step of the iteration, we take the EDCL parameter estimates we obtained in the $(k-1)-$th step $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}^{EDCL,(k-1)}, \widehat{\gamma}^{EDCL,(k-1)}, \widehat{\mu}^{EDCL,(k-1)})$ and use these parameters to estimate the RBNS component $(\widehat{X}_{ij}^{rbns,(k-1)})$, and IBNR component $(\widehat{X}_{ij}^{ibnr,(k-1)})$, using (4) and (5).

Then, we calculate the $k-$th RBNS preserving components

$$\widehat{X}_{ij}^{rbns.pres,(k)} = \frac{\sum_{j=0}^{m-i} I_{ij} - \sum_{j=0}^{m-i} X_{ij}}{\sum_{j \in \mathcal{J}(i)} \widehat{X}_{ij}^{rbns,(k-1)}} \widehat{X}_{ij}^{rbns,(k-1)}.$$

Now, we are able to define the $k-$th preliminary square $(S_{ij}^{(k)})$ as

$$S_{ij}^{(k)} = \begin{cases} X_{ij}, & \text{if } (i,j) \in \mathcal{I}, \\ \widehat{X}_{ij}^{rbns.pres,(k)} + \widehat{X}_{ij}^{ibnr,(k-1)}, & \text{if } (i,j) \in \mathcal{J}. \end{cases}$$

Since $(S_{ij}^{(k)})$ has an approximately multiplicative structure, we use the CLM idea to estimate $\widetilde{\alpha}_i$ and $\widetilde{\beta}_j$. We define

$$\widehat{\widetilde{\alpha}}_i^{EDCL,(k)} = \sum_{j=0}^{m-1} S_{ij}^{(k)}, \qquad \widehat{\widetilde{\beta}}_j^{EDCL,(k)} = \frac{\sum_{i=1}^{m} S_{ij}^{(k)}}{\sum_{(i,j)\in\mathcal{I}\cup\mathcal{J}} S_{ij}^{(k)}}.$$

Exactly as in the previous sections, we can now apply (9) and (10) to derive the EDCL parameters $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}^{EDCL,(k)}, \widehat{\gamma}^{EDCL,(k)}, \widehat{\mu}^{EDCL,(k)})$.

After iterating until convergence, we derive the final EDCL parameters denoted by $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}^{EDCL}, \widehat{\gamma}^{EDCL}, \widehat{\mu}^{EDCL})$.

4.6 The PEDCL method

In this section, we introduce the RBNS-preserving expert double chain ladder (PEDCL). As the name suggests, this method builds on the PDCL method by using the EDCL parameters $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}^{EDCL}, \widehat{\gamma}^{EDCL}, \widehat{\mu}^{EDCL})$. As mentioned in the previous section, EDCL does not preserve the RBNS case estimates. The reason was mentioned in the beginning of the previous section and has a parallelism to the two different RBNS estimates in (3) and (4).
If as in PDCL, one decides not to change the case estimates, one can thus replace the RBNS part by the RBNS derived from the case estimates. As this information is only available per underwriting year, it can be incorporated by changing the inflation parameter accordingly. More precisely, we define a new scaled inflation factor estimate as

$$\widehat{\gamma}^{PEDCL} = \frac{\sum_{j=0}^{m-i} I_{ij} - \sum_{j=0}^{m-i} X_{ij}}{\widehat{X}_{ij}^{rbns}},$$

where $\widehat{X}_{ij}^{rbns}$ is calculated with the parameters $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}^{EDCL}, \widehat{\gamma}^{EDCL}, \widehat{\mu}^{EDCL})$ using (4). This new inflation parameter $\widehat{\gamma}^{PEDCL}$ is used to replace $\widehat{\gamma}^{EDCL}$ in the parameter set $(\widehat{\alpha}, \widehat{\beta}, \widehat{\pi}^{EDCL}, \widehat{\gamma}^{EDCL}, \widehat{\mu}^{EDCL})$ when calculating only the RBNS part which preserves the case estimates. Therefore, in contrast to the previous methods, PEDCL possesses two inflation parameters. Firstly, $\widehat{\gamma}^{PEDCL}$ for the RBNS part, that is when calculating (4) and secondly $\widehat{\gamma}^{EDCL}$ for the IBNR part, that is when calculating (5). Note that since we are just using the EDCL parameters to calculate the IBNR part, the IBNR reserve of the PEDCL method is exactly the same as the IBNR reserve of the EDCL method.

5 Real Data Application

In this section, we apply the methods to a data set obtained from a UK motor business. The data is also available via the DCL R package (Martínez-Miranda

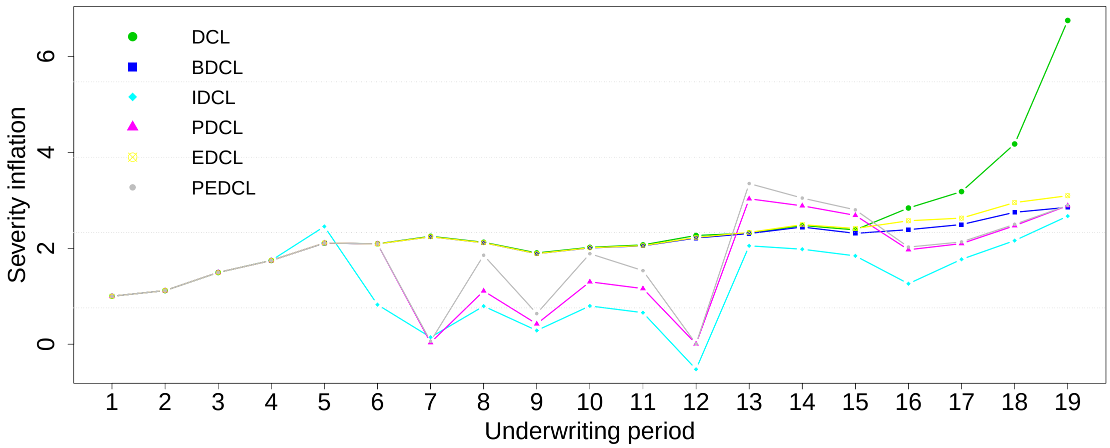

Figure 3: Plot of severity inflation estimates.

et al., 2013c). Figure 3 shows a plot of the six severity inflation parameters derived by DCL, BDCL, IDCL, PDCL, EDCL and PEDCL. There were some rather rough corrections going on for IDCL,PDCL and PEDCL in the first five years. These five years represent less than 0.1% of the total loss liabilities estimates (cf. Table 2-4) and they are not really important. We have therefore taken the first five years out for these three estimation methods. Otherwise, these unimportant first five years would have dominated the graph and perhaps have confused the reader. Figure 3 shows that IDCL,PDCL and PEDCL still are quite volatile with a tendency to have lower severity inflations than DCL, BDCL and EDCL. This is because IDCL,PDCL and PEDCL adjust to high or low RBNS's in certain years, while in particular BDCL and EDCL take a more balanced point of view. From a modeling perspective the severity inflations resulting from BDCL and EDCL seem more attractive, because they are more stable and therefore seem more realistic. This graphs illustrate very well why the PEDCL method uses the EDCL parameters while estimating the IBNR reserve. These parameters seem more realistic than the RBNS-preserving parameters. However, the RBNS-preserving parameters might be more realistic when estimating the somewhat realised RBNS reserve. If one believes the RBNS estimates are of good quality or are of the best possible quality one can do with the data, then one should of course use a method preserving these RBNS estimates such as PDCL and PEDCL. However, for the IBNR's it is another matter. There is no expert knowledge available for the IBNR estimates and therefore one should use the methodology with the most credible parameters. We believe that EDCL is the method with the most credible parameters and PEDCL is therefore - in our opinion - the optimal methodology if one wishes to preserve the RBNS estimates. If one is looking for a more balanced view, where the RBNS's are allowed to impact the parameters, but where the observed data also should play a role in some sort of validation of the RBNS as data, then one should use the EDCL method. We believe that the DCL method, the EDCL method and the PEDCL method are the three best methods to consider for practising actuaries.

To visualise these arguments, Table 2 shows the outstanding loss liabilities per underwriting year for all six methods mentioned in chapter 4 as well as the classical chain ladder method. Furthermore, the splitted values of RBNS and IBNR per underwriting year are illustrated in Table 3 and Table 4.

i	CLM	DCL	BDCL	IDCL	PDCL	EDCL	PEDCL
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0005	0.0003	0.0005	0.0000	0.0000	0.0000
3	0.0000	0.0001	-0.0003	0.0001	0.0040	0.0001	0.0040
4	0.0000	0.0007	-0.0014	0.0007	-0.0095	0.0003	-0.0095
5	0.0173	0.0039	0.0151	0.0045	0.0365	0.0076	0.0365
6	0.0346	0.0309	0.0313	0.0122	0.0054	0.0135	0.0068
7	0.1381	0.1407	0.1408	0.0090	0.0014	0.0541	0.0040
8	0.2449	0.2485	0.2483	0.0927	0.0949	0.1102	0.0972
9	0.3522	0.3582	0.3563	0.0536	0.0591	0.1774	0.0643
10	0.3943	0.3818	0.3800	0.1507	0.1981	0.2148	0.2024
11	0.5524	0.5246	0.5206	0.1664	0.2515	0.3415	0.2595
12	0.6839	0.6309	0.6169	-0.1458	0.0015	0.3415	0.0300
13	1.0504	0.9764	0.9733	0.8648	1.2718	0.8896	1.2576
14	2.5361	2.5483	2.5164	2.0388	2.8302	2.3205	2.8153
15	5.7370	5.4483	5.2846	4.2095	6.2887	5.4225	6.2721
16	14.0889	15.4373	12.9824	6.8542	9.4207	11.9785	9.4735
17	21.0057	21.7407	17.0455	12.0924	13.4243	16.5608	13.5068
18	44.6877	44.4580	29.2840	23.0002	25.9634	30.6858	26.1851
19	98.9723	98.9722	41.8444	39.1522	42.1009	45.0575	42.8640
SUM	190.4957	191.9021	112.2385	88.5565	101.9427	114.3021	103.0696

Table 2: Outstanding loss liabilities per underwriting year in million

i	CLM	DCL	BDCL	IDCL	PDCL	EDCL	PEDCL
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0005	0.0003	0.0005	0.0000	0.0000	0.0000
3	0.0000	0.0001	-0.0003	0.0001	0.0040	0.0001	0.0040
4	0.0000	0.0007	-0.0014	0.0007	-0.0095	0.0003	-0.0095
5	0.0173	0.0039	0.0151	0.0045	0.0365	0.0076	0.0365
6	0.0329	0.0291	0.0295	0.0115	0.0050	0.0118	0.0050
7	0.1355	0.1382	0.1382	0.0088	0.0014	0.0516	0.0014
8	0.2399	0.2436	0.2434	0.0909	0.0923	0.1053	0.0923
9	0.3455	0.3515	0.3496	0.0526	0.0576	0.1707	0.0576
10	0.3822	0.3697	0.3680	0.1459	0.1903	0.2027	0.1903
11	0.5339	0.5062	0.5023	0.1606	0.2411	0.3231	0.2411
12	0.6550	0.6020	0.5887	-0.1392	0.0014	0.4387	0.0014
13	1.0034	0.9294	0.9265	0.8231	1.2101	0.8420	1.2101
14	2.4415	2.4537	2.4229	1.9631	2.7197	2.2249	2.7197
15	5.5906	5.3020	5.1427	4.0964	6.1235	5.2739	6.1235
16	13.8419	15.1902	12.7746	6.7444	9.2492	11.7542	9.2492
17	20.5131	21.2482	16.6593	11.8184	13.0995	16.1534	13.0995
18	42.7693	42.5397	28.0204	22.0078	24.8281	29.3288	24.8281
19	74.0936	74.0942	31.3260	29.3108	31.4544	33.6479	31.4544
SUM	162.5956	164.0027	99.5059	77.1009	89.3045	100.5369	89.3045

Table 3: RBNS per underwriting year in million

i	CLM	DCL	BDCL	IDCL	PDCL	EDCL	PEDCL
1-5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0018	0.0018	0.0018	0.0007	0.0004	0.0018	0.0018
7	0.0026	0.0026	0.0026	0.0002	0.0000	0.0026	0.0026
8	0.0049	0.0049	0.0049	0.0018	0.0026	0.0049	0.0049
9	0.0067	0.0067	0.0067	0.0010	0.0015	0.0067	0.0067
10	0.0121	0.0121	0.0120	0.0048	0.0078	0.0121	0.0121
11	0.0184	0.0184	0.0183	0.0058	0.0103	0.0184	0.0184
12	0.0289	0.0289	0.0282	-0.0067	0.0001	0.0285	0.0285
13	0.0471	0.0471	0.0469	0.0417	0.0617	0.0476	0.0476
14	0.0946	0.0946	0.0934	0.0757	0.1105	0.0957	0.0957
15	0.1464	0.1464	0.1419	0.1131	0.1652	0.1486	0.1486
16	0.2470	0.2471	0.2077	0.1097	0.1715	0.2243	0.2243
17	0.4926	0.4925	0.3862	0.2740	0.3248	0.4073	0.4073
18	1.9183	1.9183	1.2636	0.9924	1.1353	1.3570	1.3570
19	24.8787	24.8779	10.5185	9.8414	10.6465	11.4096	11.4096
SUM	27.9001	27.8993	99.5059	11.4556	12.6382	13.7651	13.7651

Table 4: IBNR per underwriting year in million

6 Model validation

This section describes the validation process for the six methods DCL, BDCL, IDCL, PDCL, EDCL and IPDCL discussed in Section 4.

The following validation was introduced in Agbeko et al. (2015). To our knowledge this is the only developed method which is able to validate reserving estimates based on incurred data with estimates based on paid data. That is to validate DCL against IDCL in our terminology. However, this validation methodology is sufficiently general to allow all these six procedures to be validated against each other. The validation process builds on back testing and is based on the fact that all introduced DCL methods provide reserve estimates by predicting into the same paid triangle, X. Therefore, all methods can be compared on the same scale so to speak. Below, we have omitted the most recent calendar year and the four most recent calendar years, respectively (in all three available triangles). Therefore, since our dataset consists of $m = 19$ years, there are 18 and 60 cells, respectively, to be compared with the true values.

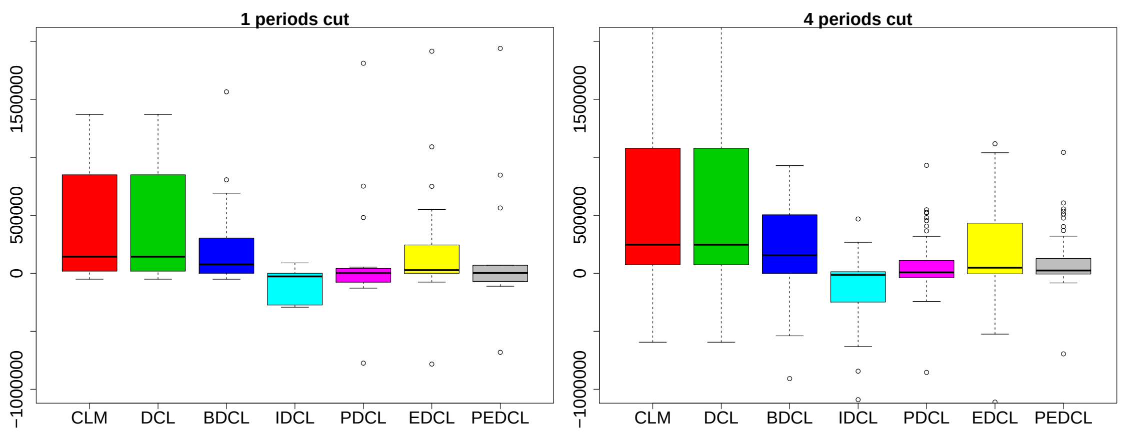

Figure 4: Box plot of the cell errors

Figure 4 shows two box plots of the respectively 18 and 60 errors calculated by taking the difference between estimated and true values. One conclusion is that DCL and CLM seem to be inferior to the five methods that take advantage of expert knowledge. Between these five methods, it seems that BDCL and EDCL have similar performance with a slight advantage to the new EDCL method. Also PDCL and PEDCL have similar performance, which was to be expected because they only differ in the IBNR reserves while having identical RBNS reserves. The omnipresent IDCL method does not provide convincing performance in this data example. Earlier studies based on a number of data sets have shown that IDCL sometimes have very good performance, but equally often fails badly. Of the three methods available at the time, DCL, BDCL and IDCL, only the BDCL method had a stable performance. Sometimes DCL was winning convincingly, other times IDCL was the winner. In the long run, however, the eternal runner-up was the BDCL method, and most of the time it did almost as well - but not quite - as the best of the two other methods. We do, however, find EDCL a more convincing method than BDCL. While they have some similarities and while both have stable underwriting year severities as we saw in the application, EDCL seems more theoretically correct in its way of exploring the expert knowledge and we believe it will be replacing BDCL in the long run. Based on long-term considerations, we think DCL, EDCL and PEDCL should be sufficient in the practical actuaries tool box. BDCL, IDCL and PDCL are - in our view - less convincing methods.

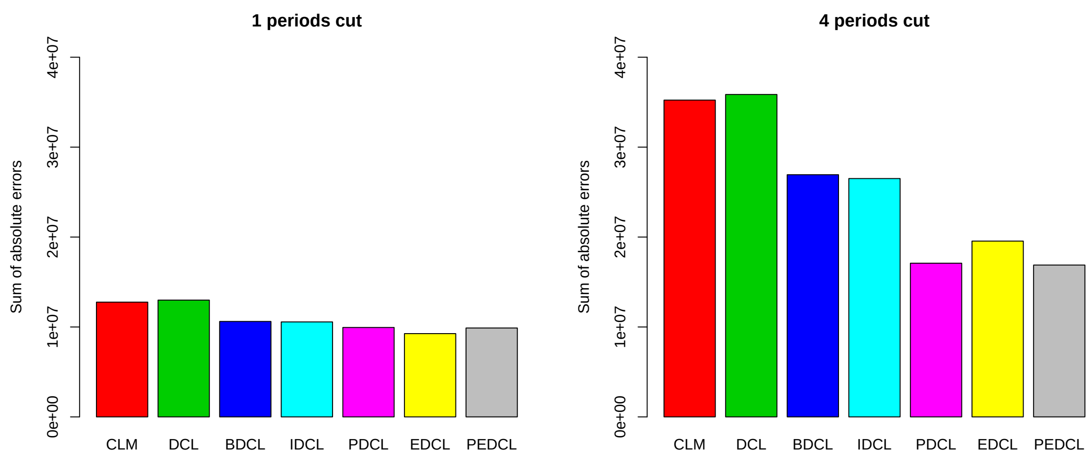

Figure 5: Bar plot for the sum of absolute cell errors

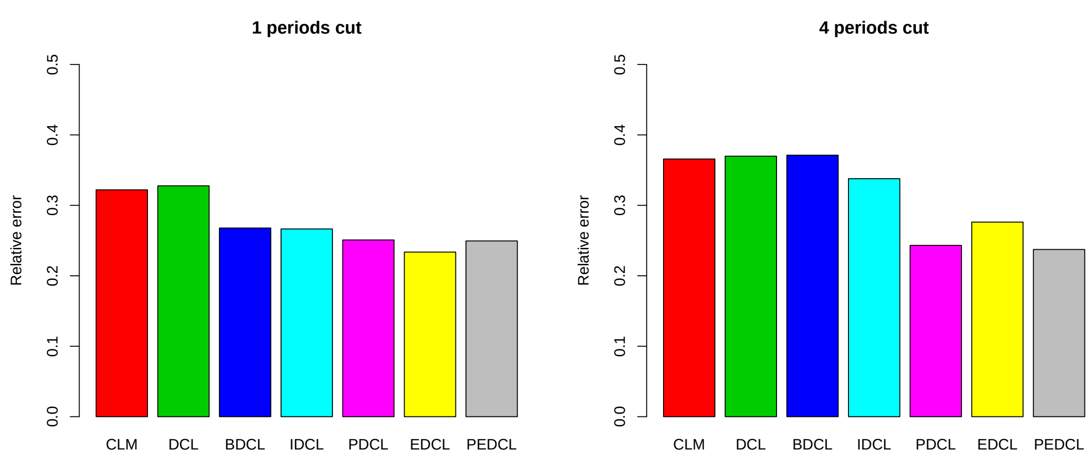

Figure 6: Bar plot for the relative errors

In the top panels of Figure 5, we have plotted the sum of the absolute cell errors (ℓ_1 error). That is

$$\text{Sum of absolute cell errors} = \sum_{(i,j)\in\mathcal{B}} |\widehat{X}_{ij} - X_{ij}|,$$

$$\mathcal{B} = \{(i,j)|\ i = 2,\ldots,m-c;\ j = 0,\ldots,m-c-1;\ i+j = m-c+1,\ldots,m\},$$

where c is the number of recent calendar years omitted for back testing (here: 1 and 4).

The relative errors, that is

$$\frac{\text{Sum of absolute cell errors}}{\text{Sum of absolute true values}} = \frac{\sum_{(i,j)\in\mathcal{B}} |\widehat{X}_{ij} - X_{ij}|}{\sum_{(i,j)\in\mathcal{B}} |X_{ij}|},$$

is shown in the bottom panels of Figure 5. The conclusion Figure 5 is similar to the conclusion of Figure 4.

7 Conclusions

This paper has developed two new methods combining classical chain ladder methodology with expert knowledge via the double chain ladder methodology. The new EDCL introduces RBNS's as pseudo data and uses an iterative procedure to improve the originally estimated DCL parameters that did not take RBNS information into account. It shows to have very good performance. The new PEDCL method preserves RBNS estimates also after the reserve has been estimated. Validation is introduced for these two new methods and they are compared to the previous methods DCL, BDCL, IDCL and PDCL. Our conclusion is that while DCL always will be some kind of benchmark in the actuaries tool box, then EDCL and PEDCL seem to have sufficient quality to replace BDCL, IDCL and PDCL. That also means that EDCL and PEDCL seem to have sufficient quality to replace incurred chain ladder as the actuaries preferred method of incorporating incurred data expert knowledge. This is important, because most reserves in the actuarial practise use incurred chain ladder as the basis of estimation. This incurred chain ladder might be manually manipulated according to expert knowledge and the values of paid chain ladder. However, it is the most common basis of estimation in actuarial practise. Only the future will be able to show whether EDCL and PEDCL or other similar innovations will be able to replace actuaries habit of using the - in our view outdated - incurred chain ladder approach.

References

Abbasi, B. and Guillén, M. (2013). Bootstrap control charts in monitoring value at risk in insurance. *Expert Systems with Applications*, **40**(15), 6125–6135.

Agbeko, T., Hiabu, M., Martínez-Miranda, M. D., Nielsen, J. P., Verrall, R. (2015). Validating the Double Chain Ladder Stochastic Claims Reserving Model. *Variance* **8**(2), 1–23.

Alai, D., Merz, M. and Wüthrich, M. V. (2009). Mean square error of prediction in the Bornhuetter-Ferguson claims reserving method. *Ann. Actuar. Sci.* **4**(1), 7–31.

Alai, D., Merz, M. and Wüthrich, M. V. (2010). Prediction uncertainty in the Bornhuetter-Ferguson claims reserving method: revisited. *Ann. Actuar. Sci.* **5**(1), 7–17.

Antonio, K. and Plat, R. (2014). Micro-level stochastic loss reserving in general insurance. *Scandinavian Actuarial Journal* **2014**(7), 649–669.

Belles-Sampera, J., Guillén, M. and Santolino, M. (2014). Beyond value-at-risk in finance and insurance: GlueVaR distortion risk measures. *Risk Analysis* **34**(1), 121–134.

Bornhuetter, R. L. and Ferguson, R. E. (1972). The actuary and IBNR. *Casualty Actuarial Society Proceedings*, 181–195.

Guelman, L. and Guillén, M. (2014). A causal inference approach to measure price elasticity in Automobile Insurance. *Expert Systems with Applications* **41**(2), 387–396.

Guillén, M., Nielsen, J.P., Scheike, T.H. and Pérez-Marín, A.M. (2012). Time-varying effects in the analysis of customer loyalty: A case study in insurance. *Expert Systems with Applications* **39**(3), 3351–3358.

Halliwell, L.J. (1997). Cojoint prediction of paid and incurred losses. *CAS Forum Summer.* 241–379.

Halliwell, L.J. (2009). Modeling paid and incurred losses together. *CAS E-Forum Spring.*

Happ, S., Merz, M. and Wüthrich, M.V. (2012). Claims development result in the paid-incurred chain reserving method. Insurance: Mathematics and Economics, **51**(1), 66–72.

Happ, S. and Wüthrich, M.V. (2013). Paid-incurred chain reserving method with dependence modeling. ASTIN Bulletin, **43**(1), 1–20.

Hiabu, M., Margraf, C., Martínez-Miranda, M. D., Nielsen, J. P. (2016). The link between classical reserving and granular reserving through double chain ladder and its extensions. *British Actuarial Journal.* **21**, 18–33.

Kaishev, V.K., Nielsen, J.P., and Thuring, F. (2013). Optimal customer selection for cross-selling of financial services products. *Expert Systems with Applications* **40**(5), 1748–1757.

Kuang, D., Nielsen, B. and Nielsen, J. P. (2009). Chain Ladder as Maximum Likelihood Revisited. *Ann. Actuar. Sci.* **4**(1), 105–121.

Mack, T. (1991) A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *Astin Bull.* **21**(1), 93–109.

Mack, T. (2008) The prediction error of Bornhuetter-Ferguson. *Astin Bull.* **38**(1), 87–103.

Martínez-Miranda, M.D., Nielsen, B., Nielsen, J.P. and Verrall, R. (2011) Cash flow simulation for a model of outstanding liabilities based on claim amounts and claim numbers. *Astin Bull.*, **41**(1), 107–129.

Martínez-Miranda, M. D., Nielsen, J. P. and Verrall, R. (2012). Double Chain Ladder. *Astin Bull.* **42**(1), 59–76.

Martínez-Miranda, M. D., Nielsen, J. P Sperlich, S. and Verrall, R. (2013a). Continuous Chain Ladder: Reformulating and generalising a classical insurance problem. *Expert Systems with Applications* **40**, 5588–5603.

Martínez-Miranda, M. D., Nielsen, J. P and Verrall, R. (2013b). Double Chain Ladder and Bornhutter-Ferguson. *N. Am. Actuar. J.* **17**(2), 101–113.

Martínez-Miranda, M. D., Nielsen, J. P and Verrall, R. (2013c). R-package "DCL": Claims Reserving under the Double Chain Ladder Model (v. 0.1.0, 25 October 2013). `http://cran.r-project.org/web/packages/DCL/index.html`.

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

Nielsen, J. P. (2016). The link between classical reserving and granular reserving through double chain ladder and its extensions. Transcripts of the London discussion. *British Actuarial Journal.* **21**, 1–17.

Pigeon, M., Antonio, K., Denuit, M. (2014). Individual loss reserving using paid–incurred data. *Insurance: Mathematics and Economics.* **58**, 121–131.

Posthuma, B., Cator, E.A., Veerkamp, W., Zwet, van E.W. (2008). Combined analysis of paid and incurred losses. *CAS E-Forum Fall.* 272–293.

Quarg, G. and Mack, T. (2004). Munich chain ladder. *Blätter der DGVFM* **26**(4), 597–630.

Schmidt, K. D. and Zocher, M. (2008). The Bornhuetter-Ferguson Principle. *Variance* **2**(1), 85–110.

Venter, G. G. (2008). Distribution and value of reserves using paid and incurred triangles. *CAS E-Forum Fall.* 348–375.

Verrall, R. (1991). Chain ladder and Maximum Likelihood. *Journal of the Institute of Actuaries* **118**, 489–499.

Verrall, R. (2004). Stochastic models for the Bornhuetter-Ferguson technique. *N. Am. Actuar. J.* **8**(3), 67–89.

Verrall, R., Nielsen, J. P. and Jessen, A. (2010). Including Count Data in Claims Reserving. *Astin Bull.* **40**(2), 871–887.