Stockholms universitet

Reserving in Presence of Backlog Claims

Philip Åberg

Masteruppsats i försäkringsmatematik
Master Thesis in Actuarial Mathematics

Masteruppsats 2026:16
Försäkringsmatematik
Juni 2026

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2026:16**
http://www.math.su.se

Reserving in Presence of Backlog Claims

Philip Åberg*

June 2026

Abstract

This thesis studies insurance claims reserving in the presence of unprocessed backlog claims. Through a simulation study, we evaluate how well different reserving methods applied to recovered reported claims perform compared to chain-ladder on processed claims and oracle reported claims, given different capacities of the claims handling unit, and how much information we obtain from a claim when processed. Reported claims are recovered through penalized least-squares optimization. Key findings are that correctly estimating the maximum development period and using a non-zero regularization parameter are both critical for reserving performance. In addition, a generalized credibility predictor of outstanding claims adapting to the backlog yields results comparable to chain-ladder on oracle reported claims.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: philip.erik.aberg@gmail.com. Supervisor: Filip Lindskog.

Contents

Preface

AI statement. In this report generative AI has been used to assist writing the computational code. It has fully made the code for the plots and tables. It has also helped to reformulate sentences to be more coherent and grammatically correct. The analysis and methods used has solely been done by the author and the guidance of the supervisor.

Acknowledgments. I want to thank my supervisor Filip Lindskog for giving me quick and detailed answers during the writing of this thesis! Some honorable mentions: Siere Herbert Erik Rasmus Mikael Heimdal, Knut Salomonsson, Christoffer Hansson, Gustav Nilsson Gisleskog, my girlfriend Linnea, my pet dogs Tripple T and Tralalero tralala and lastly the older one of my sisters.

1 Introduction

Insurance claims reserving relies on the assumption that the actuary has access to reported claims data that reflects the underlying development patterns. In practice, however, claims must be processed by a claims handling unit before they enter the data system.

When processing capacity is limited, a backlog of unprocessed claims accumulates, and what the actuary observes is not reported claims but processed claims, a distorted version of the true data.

Lindskog and Wüthrich (2025) [1] formalize this problem using a first-come-first-served queueing model, where incoming reported claims share a fixed processing capacity. They show how the numbers of reported claims can be recovered from processed claims data via a penalized least squares optimization.

This thesis implements the recovery procedure found in [1], in a simulation-based setting. The focus is on comparing different reserving methods on the recovered reported claims, with the chain-ladder estimate obtained from the reported and processed data.

2 The Lindskog-Wüthrich Framework

2.1 Reported, Processed and Backlog Claims

Let $R_{i,j}$ denote the number of reported claims from occurrence period i and development period j. Occurrence period i starts at time $i-1$ and ends at time i. Periods may refer to weeks, months, years or even seconds. In our setting we assume there exists a deterministic integer J s.t $R_{i,j} = 0$ for $j > J$. This might be to some agreement in a contract or just the nature of how reportings arrive. For clarity, this J will be assumed to be unknown and estimated later.

Further let $P_{i,j}$ be the number of claims due to events at occurrence period i and processed at development period j. Let $B_{i,j}$ denote the number of backlog claims due to events during occurrence period i that are already reported, but not processed at development period j. Table 1 illustrates the structure of $R_{i,j}$ and $P_{i,j}$ observed at calendar period $t = 5$, with $J = 3$. Note that processed claims may be observed beyond development period J.

$i \backslash j$	0	1	2	3
1	$R_{1,0}$	$R_{1,1}$	$R_{1,2}$	$R_{1,3}$
2	$R_{2,0}$	$R_{2,1}$	$R_{2,2}$	$R_{2,3}$
3	$R_{3,0}$	$R_{3,1}$	$R_{3,2}$	
4	$R_{4,0}$	$R_{4,1}$		
5	$R_{5,0}$			

$i \backslash j$	0	1	2	3	4
1	$P_{1,0}$	$P_{1,1}$	$P_{1,2}$	$P_{1,3}$	$P_{1,4}$
2	$P_{2,0}$	$P_{2,1}$	$P_{2,2}$	$P_{2,3}$	
3	$P_{3,0}$	$P_{3,1}$	$P_{3,2}$		
4	$P_{4,0}$	$P_{4,1}$			
5	$P_{5,0}$				

Table 1: Reported claims $R_{i,j}$ (left) and processed claims $P_{i,j}$ (right), observed up to calendar period $t = 5$. Empty cells correspond to unobserved future periods.

We have that the processing unit can process up to C_t each calendar period t and we will assume it is constant for each period. The total number of reported, processed and backlog claims, respectively, during calendar period t is

$$R_t := \sum_{i=t-J}^{t} R_{i,t-i} = \sum_{j=0}^{J} R_{t-j,j},$$
$$P_t := \sum_{i=i_0}^{t} P_{i,t-i} = \sum_{j=0}^{t-i_0} P_{t-j,j},$$
$$B_t := \sum_{i=i_0}^{t} B_{i,t-i} = \sum_{j=0}^{t-i_0} B_{t-j,j}.$$

Assumption 1. *The first assumption that we make is that the stochastic system* $\{(B_{i,j}, R_{i,j}, P_{i,j}, C_{i+j}) : i \geq i_0,\ j \geq 0\}$ *satisfies the following 4 conditions*

$$R_{i,j} \geq 0, \quad B_{i,j} \geq 0, \quad B_{i,0} = 0, \quad P_{i,j} \geq 0 \quad \text{for all } i, j, \tag{1}$$
$$\sum_{j=0}^{\infty} R_{i,j} = \sum_{j=0}^{\infty} P_{i,j} \quad \text{for all } i, \tag{2}$$
$$B_{i,j+1} = B_{i,j} + R_{i,j} - P_{i,j} \quad \text{for all } i, j, \tag{3}$$
$$P_t = \min(B_t + R_t,\ C_t), \tag{4}$$

Some intuitive motivation for these assumptions in order would be that we start with 0 backlog and number of claims, reported, processed or in the backlog is always positive; all reportings eventually gets processed; backlog for some occurrence period, next development period is the current plus the the reported less the processed; and lastly we can at most process C_t claims per calendar period. The event of sufficient capacity to clear all reported and backlog claims at period t will be denoted

$$\mathrm{SC}_t := \{B_t + R_t \leq C_t\}.$$

2.2 Claims Processing Procedure

The procedure we choose to process the reported claims is by first come first served (FCFS). This will imply that we clear the backlog to our capability before processing the newly reported claims. Let

$$\mathrm{SCB}_t := \{B_t \leq C_t\},$$

be the event that we have sufficient capacity to clear the backlog, at time t. At time t, what we process from the backlog together with what we process from the newly reported claims is the total amount we process

$$P_{i,t-i} = P^B_{i,t-i} + P^R_{i,t-i}. \tag{5}$$

If we define $\mathcal{F}_t := \sigma(R_s, B_s, P_s : s \leq t)$ and $\mathcal{G}_t := \sigma(R_{i,j}, B_{i,j} : i + j \leq t, j \geq 0)$. An assumption we make is that if at time t we have enough capacity to process the entire backlog, all backlog claims will be processed, and if not backlog claims would be processed independently with probability $\frac{C_t}{B_t}$ i.e., the proportion of the capacity to the size of the backlog. Or stated as given $\sigma\{\mathcal{F}_t, \mathcal{G}_t\}$ is that at time t, the claims in the backlog will be processed with conditional probability

$$\mathbb{1}_{\mathrm{SCB}_t} + \frac{C_t}{B_t}\mathbb{1}_{\mathrm{SCB}_t^c} = \mathbb{1}_{\{B_t \leq P_t\}} + \frac{P_t}{B_t}\mathbb{1}_{\{B_t > P_t\}}. \tag{6}$$

Likewise for the reported claims, if there is sufficient to process all the backlog and all newly reported claims, then the probability for an individual claim to be processed is one, otherwise it will be processed with the probability of the proportion of the remaining capacity to the number of reported claims

$$\mathbb{1}_{\mathrm{SC}_t} + \frac{C_t - B_t}{R_t}\mathbb{1}_{\mathrm{SC}_t^c \cap \mathrm{SCB}_t} = \frac{P_t - B_t}{R_t}\mathbb{1}_{\{B_t \leq P_t\}}. \tag{7}$$

In conclusion $P^B_{i,t-i}$ and $P^R_{i,t-i}$ given $\sigma\{\mathcal{F}_t, \mathcal{G}_t\}$ are binomial distributed

$$P^B_{i,t-i} \mid \sigma\{\mathcal{F}_t, \mathcal{G}_t\} \sim Bin\left(B_{i,t-i}, \mathbb{1}_{\{B_t \leq P_t\}} + \frac{P_t}{B_t}\mathbb{1}_{\{B_t > P_t\}}\right), \tag{8}$$

$$P^R_{i,t-i} \mid \sigma\{\mathcal{F}_t, \mathcal{G}_t\} \sim Bin\left(R_{i,t-i}, \frac{P_t - B_t}{R_t}\mathbb{1}_{\{B_t \leq P_t\}}\right). \tag{9}$$

Then equation (5), (8) and (9) summarizes how the procedure for processing claims is defined. Using (5) we obtain

$$\mathbb{E}\left[P_{i,t-i} \mid \sigma\{\mathcal{F}_t, \mathcal{G}_t\}\right] = B_{i,t-i}\left(\mathbb{1}_{\mathrm{SCB}_t} + \frac{P_t}{B_t}\,\mathbb{1}_{\mathrm{SCB}_t^c}\right) + R_{i,t-i}\left(\frac{P_t - B_t}{R_t}\,\mathbb{1}_{\mathrm{SCB}_t}\right). \tag{10}$$

Recall (3), by iteratively using this property we obtain

$$B_{i,t-i} = \left(\mathbb{1}_{\{t-i>0\}}\sum_{j=0}^{t-i-1}(R_{i,j} - P_{i,j})\right), \tag{11}$$

which substituted into (10) yields

$$\mathbb{E}\left[P_{i,t-i} \mid \sigma\{\mathcal{F}_t, \mathcal{G}_t\}\right] = \left(\mathbb{1}_{\{t-i>0\}}\sum_{j=0}^{t-i-1}(R_{i,j} - P_{i,j})\right)\left(\mathbb{1}_{\mathrm{SCB}_t} + \frac{P_t}{B_t}\,\mathbb{1}_{\mathrm{SCB}_t^c}\right) + R_{i,t-i}\left(\frac{P_t - B_t}{R_t}\,\mathbb{1}_{\mathrm{SCB}_t}\right). \tag{12}$$

Together with assumtion 1, we will also assume the following:

Assumption 2. *The stochastic system* $\{(B_{i,j}, R_{i,j}, P_{i,j}, C_{i+j}) : i \geq i_0,\, j \geq 0\}$ *satisfies equation* (10).

2.3 Recovering of Reported Claims from Processed Data

2.3.1 Estimation of Reported Claims

For reserving at time t an actuary would like to know

$$\{R_{i,j} : 1 \leq i \leq t, i + j \leq t\},$$

i.e. the true reportings. However the core problem is that the actuaries in practice do not know this, but instead relies on the processed data instead. Making reasonable assumptions on what the actuary has observed up until time t, we will see that the true underlying reported claims could be estimated. We assume that the actuary has observed up until time t

$$\{B_s : 1 \leq s \leq t+1\} \cup \{R_s, P_s : 1 \leq s \leq t\} \cup \{P_{i,j} : 1 \leq i \leq t, i + j \leq t\}.$$

Using (12) and $\varepsilon_{i,j} := \mathbb{E}[P_{i,j} \mid \sigma\{\mathcal{F}_{i+j}, \mathcal{G}_{i+j}\}] - P_{i,j}$ we obtain

$$\begin{aligned} P_{i,j} + \varepsilon_{i,j} = &\left(\mathbb{1}_{\{j>0\}}\sum_{k=0}^{j-1}(R_{i,k} - P_{i,k})\right)\left(\mathbb{1}_{\mathrm{SCB}_{i+j}} + \frac{P_{i+j}}{B_{i+j}}\,\mathbb{1}_{\mathrm{SCB}_{i+j}^c}\right) \\ &+ R_{i,j}\left(\frac{P_{i+j} - B_{i+j}}{R_{i+j}}\,\mathbb{1}_{\mathrm{SCB}_{i+j}}\right). \end{aligned}$$

This suggest that one way we could estimate the $R_{i,j}$'s is to minimize $\sum \varepsilon^2_{i,j}$ with the constraint that the reportings are positive and that their diagonal sum should be equal to the observed cumulative values for each period. Or alternatively we look at the Lagrange version where we penalize on the diagonal sum, which we will do. Hence we minimize

$$\begin{aligned} &\sum_{i=1}^{t}\sum_{j=0}^{t-i}\left(a_{i+j}\mathbb{1}_{\{j>0\}}\sum_{k=0}^{j-1}(r_{i,k} - P_{i,k}) + b_{i+j}r_{i,j} - P_{i,j}\right)^2 \\ &+\gamma\sum_{s=J}^{t}\left(\sum_{j=0}^{J} r_{s-j,j} - R_s\right)^2, \end{aligned} \tag{13}$$

over all non-negative integers $r_{i,j}$. We also suppose that we estimate J. The coefficients a_{i+j} and b_{i+j} are given by

$$a_{i+j} = \mathbb{1}_{\mathrm{SCB}_{i+j}} + \frac{P_{i+j}}{B_{i+j}}\,\mathbb{1}_{\mathrm{SCB}^c_{i+j}},$$
$$b_{i+j} = \frac{P_{i+j} - B_{i+j}}{R_{i+j}}\,\mathbb{1}_{\mathrm{SCB}_{i+j}}.$$

2.3.2 Matrix Formulation

It is possible to reformulate equation (13) in matrix form. Suppose that we have estimated $\hat{J}$ as an upper bound for J, and let n denote the number of occurrence periods under consideration. The unknown reportings $r_{i,j}$ are reshaped into a column vector $\mathbf{r} \in \mathbb{R}^{n(\hat{J}+1)}$ and the observed processed claims $P_{i,j}$ into a corresponding vector $\mathbf{p} \in \mathbb{R}^{n(\hat{J}+1)}$, both ordered row-wise so that occurrence period i occupies positions $(i-1)(\hat{J}+1)+1$ through $i(\hat{J}+1)$.
For each occurrence period $i \in \{1,\dots,n\}$, the matrices $M_r^{(i)}$ and $M_p^{(i)}$ are of square dimension $(\hat{J}+1)\times(\hat{J}+1)$. For each row $j \in \{0,\dots,\hat{J}\}$ the relevant calendar period is $s = i+j$. The entries are determined by a_s and b_s in the following way, where for unobserved s, $a_s = b_s = 0$.

$$M_r^{(i)} = \begin{bmatrix} b_i & 0 & 0 & \cdots & 0 \\ a_{i+1} & b_{i+1} & 0 & \cdots & 0 \\ a_{i+2} & a_{i+2} & b_{i+2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i+\hat{J}} & a_{i+\hat{J}} & a_{i+\hat{J}} & \cdots & b_{i+\hat{J}} \end{bmatrix}, \quad M_p^{(i)} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{i+1} & 1 & 0 & \cdots & 0 \\ a_{i+2} & a_{i+2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i+\hat{J}} & a_{i+\hat{J}} & a_{i+\hat{J}} & \cdots & 1 \end{bmatrix}.$$

The full matrices M_r and M_p are then block-diagonal of dimension $n(\hat{J}+1)\times n(\hat{J}+1)$, with $M_r^{(i)}$ and $M_p^{(i)}$ as the i-th diagonal block:

$$M_r = \begin{bmatrix} M_r^{(1)} & & \\ & \ddots & \\ & & M_r^{(n)} \end{bmatrix}, \quad M_p = \begin{bmatrix} M_p^{(1)} & & \\ & \ddots & \\ & & M_p^{(n)} \end{bmatrix}. \tag{14}$$

The first square error term in (13) can then be written as

$$\|M_r\mathbf{r} - M_p\mathbf{p}\|^2.$$

When $\gamma > 0$, the penalty term enforces the diagonal-sum constraints $\sum_{j=0}^{\tilde{J}} r_{s-j,\,j} = R_s$, for a separate estimate $\tilde{J} \le \hat{J}$, which is our actual estimate and not just an upper bound. The matrix A has dimension $(n-\tilde{J})\times n(\tilde{J}+1)$, with one row per observable calendar period $s \in \{\tilde{J}+1,\dots,n\}$. Row $k \in \{1,...,n-\tilde{J}\}$, corresponding to calendar period $s = \tilde{J}+k$, contains a 1 at column position $(i-1)\cdot(\hat{J}+1)+j+1$ for each pair (i,j) satisfying $i = s-j$, $0 \le j \le \tilde{J}$ and $1 \le i \le n$, and zeros elsewhere. Let $\mathbf{s} = (R_{\tilde{J}+1},..,R_n)$, we can then formulate the constraint term in (13) as

$$\gamma\|A\mathbf{r} - \mathbf{s}\|^2$$

The full penalized least-squares problem is therefore in matrix form

$$\min_{\mathbf{r}\ge 0} \|M_r\mathbf{r} - M_p\mathbf{p}\|^2 + \gamma\|A\mathbf{r} - \mathbf{s}\|^2.$$

Since it holds for arbitrary vectors that $\|u\|^2 + \|v\|^2 = \left\|\begin{bmatrix} u \\ v\end{bmatrix}\right\|^2$, where u and v are stacked, applied with $u = M_r\mathbf{r} - M_p\mathbf{p}$ and $v = \sqrt{\gamma}(A\mathbf{r} - \mathbf{s})$, this is in turn equivalent to

$$\min_{\mathbf{r}\ge 0} \|C\,\mathbf{r} - \mathbf{d}\|^2, \quad C = \begin{bmatrix} M_r \\ \sqrt{\gamma}\,A \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} M_p\,\mathbf{p} \\ \sqrt{\gamma}\,\mathbf{s} \end{bmatrix}. \tag{15}$$

3 Preliminaries

3.1 Reserving

Reserving is an essential part for insurance companies. It is a way of saying ”saving money for future claims”. There are several methods of doing this, but we will follow a method given in Neuhaus [3]. Let $\{X_j : j = 0, 1, ...\}$ denote the incremental claims from a given occurrence period, here j is the development period. Denote by D the number of development periods that have elapsed since the start of the given occurrence period at the valuation date. Then assume we have observed $\{X_j : j \leq D\}$, the task is then to predict the outstanding claims $\{X_j : j > D\}$. Let β be a priori predictor of $\sum_{j=0}^{\infty} X_j$ such that they are equal in expectation. Let π_j be the expected proportion of claims that will emerge in development period j, naturally $\pi_j \geq 0$ and $\sum_{j=0}^{\infty} \pi_j = 1$. We will assume β and $\{\pi_j : j = 0, 1, ...\}$ are estimated and fixed. Then a general linear credibility predictor of outstanding claims is

$$\overline{X}_{>D}(z) = \pi_{>D} \cdot (zb^* + (1 - z)\beta), \tag{16}$$

where

$$b^* = \frac{X_{\leq D}}{\pi_{\leq D}}.$$

A subscript with an inequality sign means that we sum over all indices j satisfying the inequality. We choose to estimate $\pi_{>D}$ by first estimating the development factors $\{f_j : j = 1, 2, ...\}$ in the same way as in the famous chain ladder (see Example 1 below for a reminder). Hence

$$\hat{\pi}_{>D} = 1 - \frac{1}{\prod_{j>D} \hat{f}_j}. \tag{17}$$

For our applications we define $\hat{f}_j = 1$ for large values of j for which we have no observations. If this is how we choose to estimate $\{\pi_j : j = 0, 1, ...\}$, then b^* is the chain ladder prediction of outstanding claims. Further $z \in [0, 1]$ is what we call a credibility factor and determines to what degree we will trust the prior estimate β and the chain ladder estimate.

Example 1. Mack's chain-ladder [2], is a well-known standard method for reserving, we briefly recall the construction of the development factors in a simple example.

Let the following matrix denote the incremental claims that we have observed

$$\begin{pmatrix} X_{10} & X_{11} & X_{12} & X_{13} \\ X_{20} & X_{21} & X_{22} & \cdot \\ X_{30} & X_{31} & \cdot & \cdot \\ X_{40} & \cdot & \cdot & \cdot \end{pmatrix},$$

and let the matrix below be the corresponding cumulative one, i.e $C_{i,j} = \sum_{k \leq j} X_{i,k}$

$$\begin{pmatrix} C_{10} & C_{11} & C_{12} & C_{13} \\ C_{20} & C_{21} & C_{22} & \cdot \\ C_{30} & C_{31} & \cdot & \cdot \\ C_{40} & \cdot & \cdot & \cdot \end{pmatrix}.$$

The development factors are then estimated by the ratio of the sum of two adjacent columns in a suitable way.

$$\hat{f}_j := \frac{\sum_{i=1}^{4-j} C_{i,j}}{\sum_{i=1}^{4-j} C_{i,j-1}}.$$

3.2 Bounded-Variable Least-Squares

Bounded-variable least-squares (BVLS) is an optimization algorithm developed by P.B Stark and R.L Parker[4], which is fully implemented SciPy Python. It solves the following problem

$$\min_{l \leq x \leq u} \|Ax - b\|_2^2. \tag{18}$$

Here $l, x, u \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and A is a m by n matrix.

BVLS is an algorithm that maintains two active sets, $\mathcal{L}$ and $\mathcal{U}$ for variables at their lower and upper bound respectively and a free set $\mathcal{F}$ for variables strictly between the bounds. The algorithm ends when the conditions in the third step are met, and when it terminates, the solution is the global minimum in the region $l \le x \le u$, according to Stark & Parker [4] (p. 2). The algorithm is as follows

1. Set $\mathcal{F} = \mathcal{U} = \emptyset, \mathcal{L} = \{1, ..., n\}$, and set every element of x to its lower bound.

2. Compute $w = A^T(b - Ax)$, the negative gradient of the objective.

3. If $\mathcal{F} = \{1, ..., n\}$, or, if $w_k \le 0$ for all $k \in \mathcal{L}$ and $w_k \ge 0$ for all $k \in \mathcal{U}$, go to step 11.

4. Find $t^* = \arg\max_{t \in \mathcal{L} \cup \mathcal{U}} s_t w_t$, where $s_t = 1$ if $t \in \mathcal{L}$ and $s_t = -1$ if $t \in \mathcal{U}$. If t^* is not unique, break ties arbitrarily.

5. Move t^* to the set $\mathcal{F}$.

6. Let b' be the data vector less the predictions of the bound variables; i.e., $b'_k = b_k - \sum_{i \in \mathcal{L} \cup \mathcal{U}} A_{ki} x_i$. Let A' be the matrix composed of those columns of A whose indices are in $\mathcal{F}$. Let k' denote the index of the column of A' corresponding to the original index $k \in \mathcal{F}$. Find
$$z = \arg\min \|A'z - b'\|_2^2.$$

7. If $l_k < z_{k'} < u_k$ for all k', set $x_k = z_{k'}$ and go to step 2.

8. Let $K^- = \{k' \in \mathcal{F} : z_{k'} < l_k\}$ and $K^+ = \{k' \in \mathcal{F} : z_{k'} > u_k\}$ be the sets of indices of components of z below and above their bounds respectively, and let $K = K^- \cup K^+$. Define
$$q' = \arg\min_{k' \in K} \alpha_{k'}, \quad \text{where} \quad \alpha_{k'} = \begin{cases} \dfrac{l_k - x_k}{z_{k'} - x_k} & k' \in K^- \\ \dfrac{u_k - x_k}{z_{k'} - x_k} & k' \in K^+ \end{cases}$$
and set $\alpha = \alpha_{q'}$.

9. Set $x_k := x_k + \alpha(z_{k'} - x_k)$ for all $k \in \mathcal{F}$.

10. Move to $\mathcal{L}$ the index of every component of x at or below its lower bound. Move to $\mathcal{U}$ the index of every component at or above its upper bound. Go to step 6.

11. Done, returns a minimizing x.

Here step 3 is the criteria for being done. It checks wether or not we gain anything for adjusting our variables. Step 4 and 5 checks which variable the gradient most suggests to move inside from the boundary. Step 6 calculates among the free variables their values that minimizes the objective. Step 8 calculates which variable that would hit the boundary first when moving from x_k to $z_{k'}$, and sets the step size parameter $\alpha \in (0, 1]$ to be the scalar that moves that variable to the boundary.

The reason for applying this algorithm to problem (15) instead of other implemented algorithms, like non-negative least-squares, which solves the case of $l = 0$ and u is unbounded, is that we in some cases want to bound certain $r_{i,j}$'s, either because we want some of them fixed, or that we want control the dimension of the output.

4 Data Generation & Recovery Method

4.1 Claim Generation

While having fixed $J = 3$ as the maximum development period, we simulated 250 occurrence periods of reported claims. i.e a 250×4 matrix R with integer values in each entry. The amount of reported claims for each pair of occurrence and development periods, $R_{i,j}$, were drawn independently from a Negative Binomial distribution with parameters (α_j, β), where $\beta = 2/1000$ and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3)/\beta = (500, 300, 150, 50)$.

It holds that $\mathbb{E}[R_{i,j}] = \mu_j = \alpha_j/\beta$, additionally

$$R_t = \sum_{j=0}^{J} R_{t-j,j} \overset{d}{=} \sum_{j=0}^{J} R_{i,j} \sim NegBin(\alpha, \beta), \quad \alpha := \sum_{j=0}^{J} \alpha_j,$$

with mean $\mathbb{E}[R_t] = \mu = \alpha/\beta = 1000$.

4.2 Queue Processing

Claims enter a FCFS queue with service capacity $c = \eta\mu$, for some fixed $\eta > 1$. To implement the queue, each claim in $R_{i,j}$ is given a label consisting of (i, j) and a random timestamp $i+j-1+U_{i,j,k}$, where $U_{i,j,k}$ is uniform on $(0, 1)$. Here $k \in \{1, ..., R_{i,j}\}$ corresponds to a claim and the timestamp is a random time during calendar period $i + j$. Further they are all sorted globally by this timestamp.

At each calendar period t, the available claims consist of the current backlog and all claims with $i+j = t$. Out of these the first c, given their timestamp, are processed and the rest goes into the backlog for the next calendar period. When a claim from occurrence period i is processed at development period j, we increase $P_{i,j}$ by 1.

4.3 Recovery Method

In our analysis we are interested in predicting the ultimate claim count for the last $n = 17$ occurrence periods. Thus following section 2.3, we construct the minimization problem (15)

$$\min_{\mathbf{r}\geq 0} \|C\,\mathbf{r} - \mathbf{d}\|^2, \quad C = \begin{bmatrix} M_r \\ \sqrt{\gamma}\,A \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} M_p\,\mathbf{p} \\ \sqrt{\gamma}\,\mathbf{s} \end{bmatrix},$$

and we minimize it using the bounded-value least-squares algorithm (BVLS).

Several approaches are made to reconstruct the reported claims, which will include varying the penalty term $\gamma \in \{0, 1, 10, 100\}$ and using different estimates $\tilde{J}$ of J in the construction of the diagonal sum matrix A. In addition when having estimated $\tilde{J}$, the optimization will be done under the assumption that $r_{i,j} = 0$ for $j \geq \tilde{J}$.

Three variants are considered regarding $\tilde{J}$. First variant is that we let $\tilde{J}$ be estimated by

$$J_P := \max\{j : \; P_{i,j} > 0, \quad i + j \leq 17\},$$

while letting $\hat{J}$ in the formulation of (14) be equal to J_P. The interpretation of this estimate is that we count have many columns that are non-zero for what we have observed so far. Secondly we estimate $\tilde{J}$ by

$$J_{P,\min} := \min_{i\,<\,n-J_P} \max\{j : P_{i,j} > 0, \quad i + j \leq 17\},$$

while letting $\hat{J} = J_P$. This estimate is used to reduce overestimation of the previous one. Lastly we assume $\tilde{J} = J = 3$, i.e that we have estimated J correctly. In this case we set $\hat{J} = \max(3, J_P)$ for the purpose of using as much information we can while ensuring the dimension of our output will not fall short of what we have assumed. i.e we do not want a 17×3 matrix of recovered reportings when the assumption is that we will have 4 columns.

The quality of the recovered reported claims will then be tested by how well they work to reserve with. That is how well they predict

$$\sum_{i,j} R_{i,j} = \sum_{i=1}^{17} \sum_{j=0}^{J} R_{i,j}. \tag{19}$$

5 Chain-Ladder Applied to Recovered Numbers of Claims

All results are based on $N = 10,000$ independent replications per parameter configuration. The capacity ratio takes values $\eta \in \{1.10, 1.20, 1.40\}$ and regularization $\gamma \in \{0, 1, 10, 100\}$. For each simulation, we generate 250 occurrence periods of reported claims, such that B_t reaches its stationary distribution, which exists, for all capacity ratios of interest, [1] (p. 31). From this data, we obtain the processed claims $P_{i,j}$ and the cumulative backlog claims according to 4.2. Then the data used is R_i, $P_{i,j}$ and B_i, for $i \in \{234, 235, ..., 250\}$ i.e the observations from last 17 observed occurrence periods. For the ease of notation we write for $i \in \{1, 2, ..., 17\}$ instead.

As stated in section 4.3, we recover $\hat{R}_{i,j}$ with three different estimates for J for each pair (η, γ). We then apply Mack's chain-ladder to each of the cumulative recovered reported claims ($\hat{R}$), the cumulative processed claims (P) and to the cumulative true reportings (R). The later serves the purpose of having a reference of the error distribution and the RMSE. The error is defined by the difference in the (chain-ladder) prediction and the true ultimate reported claims which was simulated at the start

$$error = prediction - \sum_{i=1}^{17} \sum_{j=0}^{3} R_{i,j},$$

and the RMSE (root mean square error) is

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^{N} error_k},$$

over N predictions.

5.1 Quality by Parameter Configuration

From now on CL(P), CL(R), CL($\hat{R}, J$), CL($\hat{R}, J_P$) and CL($\hat{R}, J_{P,\min}$) means the processed claims, the true reported claims, recovered reportings using $\tilde{J} = 3$, recovered reportings using $\tilde{J} = J_P$ and recovered reportings using $\tilde{J} = J_{P,\min}$ where used to predict the ultimate claim count (19) by Mack's chain-ladder based on what we know up until time t respectively.

Below tables showcases the RMSE and mean error for the different recovery methods for each pair (η, γ). After analyzing the table we see that $\gamma = 100$ works best generally for when having estimated J correctly. The best choice of regularization parameter for when using the other two estimates of J, is $\gamma = 0$ for low capacity ratios. In Table 3 and Table 4 there are some values marked with a *, due to some instances where recovery did not do well. If we remove the two outliers when using the $J_{P,min}$ estimate the mean error for $(\eta = 1.1, \gamma = 1)$ will be 3432 and 2889 for the pair $(\eta = 1.1, \gamma = 100)$, the change in RMSE will be from 336M to 60k and 2.48M to 53k respectively.

Table 2: Mean Error and RMSE - CL(P) and CL(R) reference

	$\eta = 1.1$	$\eta = 1.2$	$\eta = 1.4$
Mean Error			
CL(P)	8018	5243	281
CL(R)	42	42	42
RMSE			
CL(P)	138580	141690	9929
CL(R)	902	902	902

Table 3: Mean Error - Recovery Methods

η	Method	$\gamma = 0$	$\gamma = 1.0$	$\gamma = 10$	$\gamma = 100$
	$\text{CL}(\hat{R}, J)$	70	1017	653	596
1.1	$\text{CL}(\hat{R}, J_P)$	1791	5874	5446	5545
	$\text{CL}(\hat{R}, J_{P,\min})$	1207	37575*	3454	28119*
	$\text{CL}(\hat{R}, J)$	233	375	190	156
1.2	$\text{CL}(\hat{R}, J_P)$	538	858	739	719
	$\text{CL}(\hat{R}, J_{P,\min})$	-182	318	613	729
	$\text{CL}(\hat{R}, J)$	10	95	54	45
1.4	$\text{CL}(\hat{R}, J_P)$	72	125	84	77
	$\text{CL}(\hat{R}, J_{P,\min})$	-817	-381	-69	30

Table 4: RMSE - Recovery Methods

η	Method	$\gamma = 0$	$\gamma = 1.0$	$\gamma = 10$	$\gamma = 100$
	$\text{CL}(\hat{R}, J)$	6957	2487	2042	1994
1.1	$\text{CL}(\hat{R}, J_P)$	8385	79957	78805	79860
	$\text{CL}(\hat{R}, J_{P,\min})$	8321	3361717*	70344	2476090*
	$\text{CL}(\hat{R}, J)$	4417	1467	1199	1170
1.2	$\text{CL}(\hat{R}, J_P)$	4940	8938	9511	9501
	$\text{CL}(\hat{R}, J_{P,\min})$	4924	8177	29847	37811
	$\text{CL}(\hat{R}, J)$	1316	964	924	919
1.4	$\text{CL}(\hat{R}, J_P)$	2410	1077	1027	1020
	$\text{CL}(\hat{R}, J_{P,\min})$	1738	1120	934	927

5.1.1 Distribution of Prediction Error

Below are some figures visualizing the distribution of the prediction error for the pair $(\eta, \gamma) = (1.2, 100)$. Other pairs gave similar results. Worth noting is that the lower and upper 1% has been removed for us to be able to see the distributions clearly and not as one single block/staple. The red dotted vertical line showcases the mean.

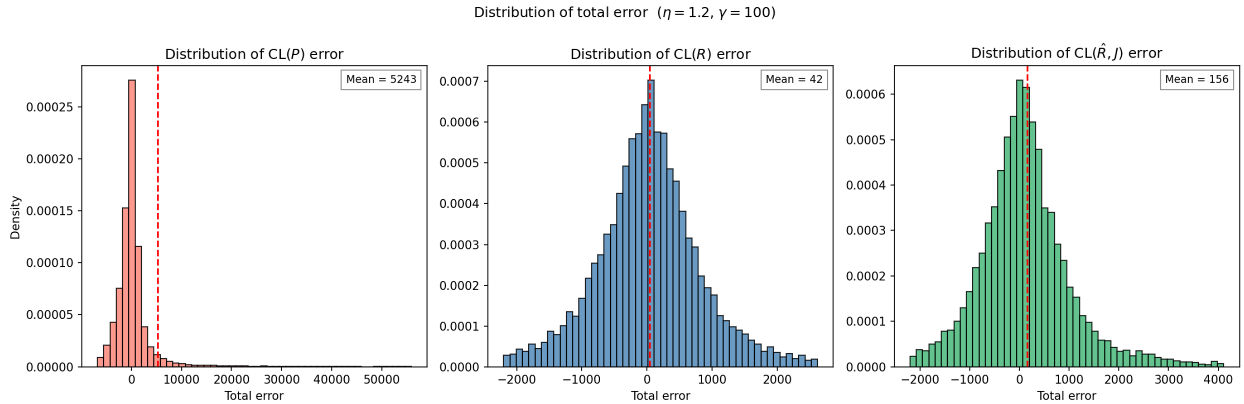

Figure 1: Distribution of the prediction error using Processed claims, True reported claims and recovered claims assuming we have estimated $J = 3$ correctly.

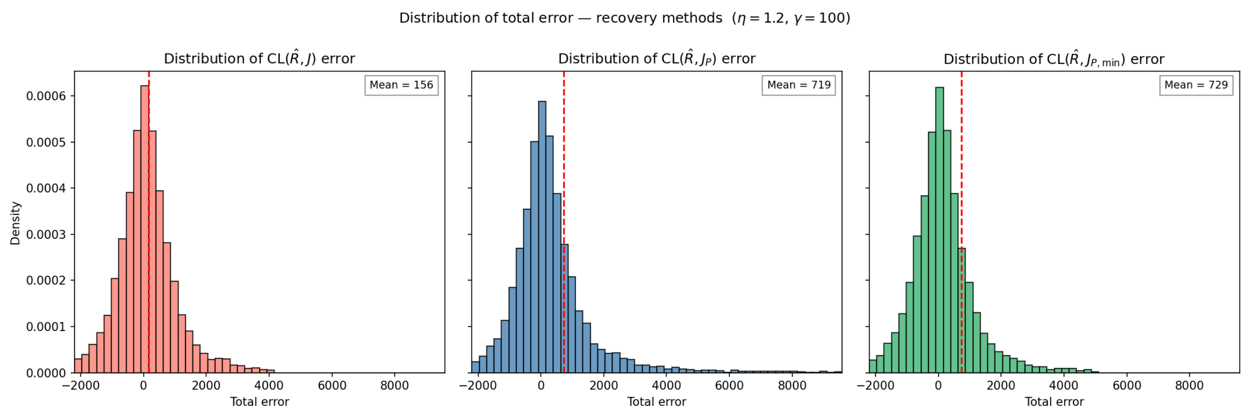

Figure 2: Distribution of the prediction error for the different methods of estimating J.

5.1.2 RMSE Comparison

These visualizations compares the RMSE for the pair $(\eta, \gamma) = (1.2, 100)$.

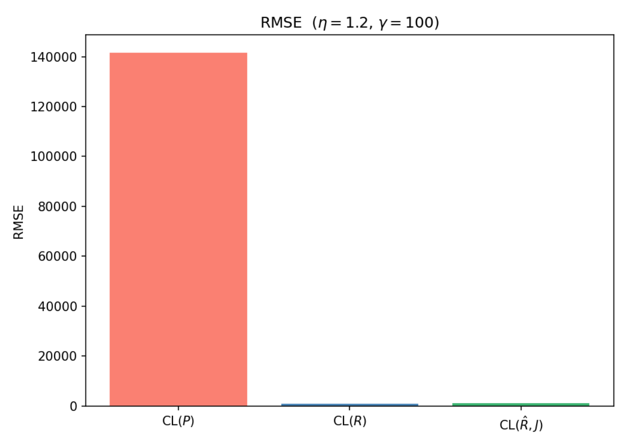

Figure 3: Root Mean Square Error using Processed claims, True reported claims and recovered claims assuming we have estimated $J = 3$ correctly.

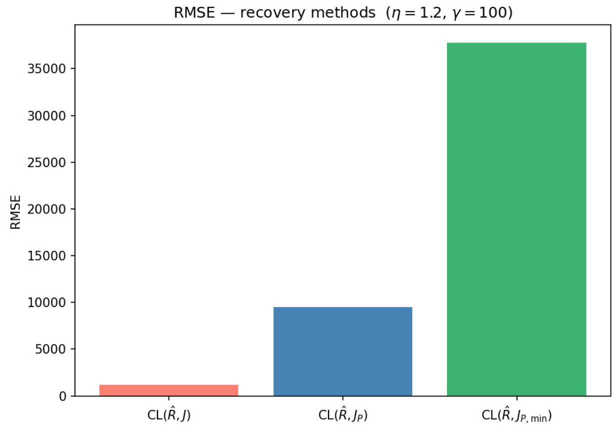

Figure 4: RMSE for the different methods of estimating J.

5.2 Occurrence Periods Used for Recovery

The setup we have now is that we are trying to determine the ultimate claim count of reported claims for the last 17 periods in our simulation. Previously we did this by only taking into account R_t, $P_{t,j}$ and B_t for $t \in \{234, 235, ..., 250\}$ into account. However, a natural question would be if the prediction of ultimate claims for the latest 17 would improve if we used more data when recovering the reportings, i.e we recreate the reportings for the latest 24 periods and then out of those we only consider the latest 17 as usual. The goal is to analyze if the removal of a systematic error is possible. Figure 5 shows for each occurrence period i the mean of

$$\sum_{j=0}^{J} \hat{R}_{i,j} - \sum_{j=0}^{J} R_{i,j}$$

across $10,000$ simulations, when recreating the reported claims for the latest 17 and 24 periods. In this simulation we have assumed to estimate $J = 3$ correctly. We have also simulated with capacity ratio $\eta = 1.1$ and used $\gamma = 100$ as the diagonal sum penalty parameter. Since given the estimate of $J = 3$ the chain-ladder prediction for an occurrence period $i \in \{1, 2, ...14\}$, again of the latest 17, will coincide with $\sum_j \hat{R}_{i,j}$, we can note that some bias will probably on average be reduced if we extend the amount of data used. This is an indication that it is more suitable to use more information when recovering the reported claims.

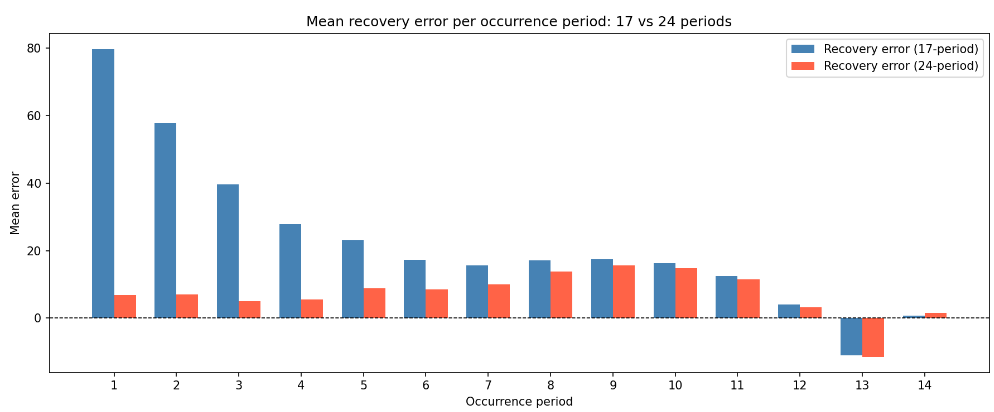

Figure 5: Mean error of the total reported claims (y-axis) by occurrence period (x-axis). 24 occurrence periods used for recovery in red and 17 occurrence periods used for recovery blue.

5.3 Estimating J

Estimating J is not trivial. An overestimation will result in more development factors used which in turn are likely to be large. In this case the CL estimate of total claims will be overestimated. Figure 6 shows the distribution of the estimates we have used and clearly they do not often agree with the true J used in simulation.

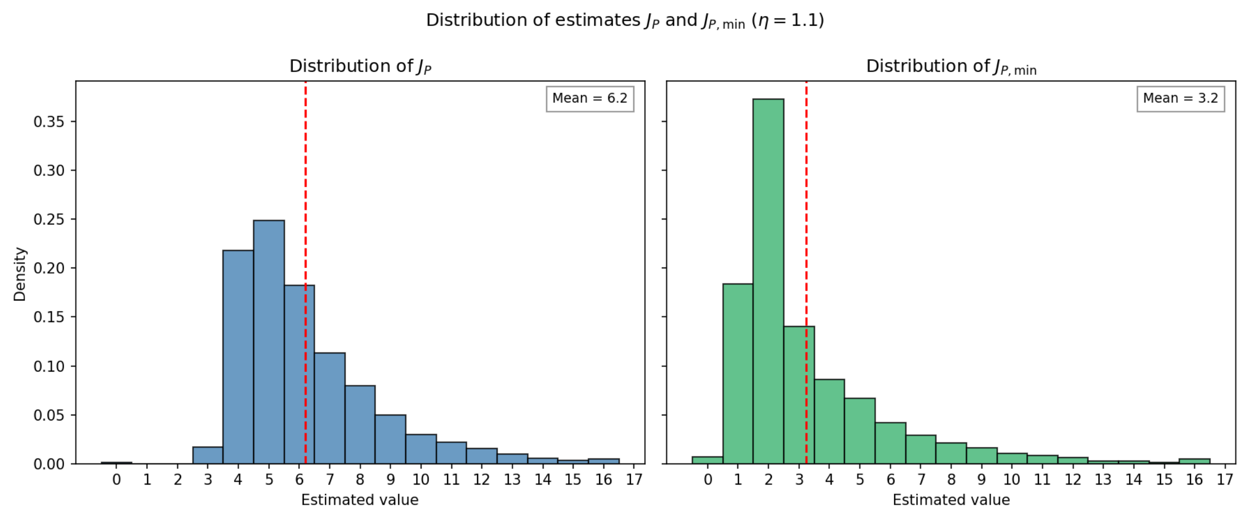

Figure 6: Distribution of J_P and $J_{P,min}$

.

6 Benktander Golden Stairs

Recall the general credibility predictor of outstanding claims from Eq. (16)

$$\overline{X}_{>D}(z) = \pi_{>D} \cdot (zb^* + (1-z)\beta)\,,$$

where

$$b^* = \frac{X_{\leq D}}{\pi_{\leq D}}$$

is the cumulative Chain Ladder estimate and β is a prior estimate of the ultimate reserve for one specific period. First thing to notice is that $z = 1, z = 0$ and $z = \pi_{\leq D}$ correspond to the Chain Ladder, Bornhuetter-Ferguson and the Benktander method respectively. In [3] they namely discusses what is optimal if we view the weight z as a function of the proportion of developed claims, i.e $z(\pi_{\leq D}, \delta)$. Here δ is just a generic parameter to make the function more concave/convex.

In our case it is reasonable to let the backlog or other related metrics in addition to the claim development to determine the shape of z. Hence we consider

$$z(\pi_{\leq D}, \delta(\kappa)) := \pi_{\leq D}^{\delta},$$

where

$$\delta = \frac{\kappa}{\overline{a}}, \qquad \overline{a} = \frac{1}{17}\sum_{s=1}^{17} a_s.$$

Clearly it means we rely more on the Bornhuetter-Ferguson (BF) estimate when we have a lot of backlogs or if we are in a more underdeveloped period. $\kappa = 0$ corresponds to chain-ladder, $\kappa = \overline{a}$ corresponds to Benktander (BK) and $\kappa \to \infty$ corresponds to BF. Figure 7 below showcases how the mean error for for the prediction of total ultimate claims, and RMSE varies across κ, were our prior guess was $\beta = \frac{1}{17}\sum_{t=1}^{17} R_t$.

10,000 independent simulations were made with capacity ratio $\eta = 1.1$ and penalty parameter $\gamma = 100$. The development factors, $\pi_{\leq D}$, were calculated in each simulation based on the development pattern of the recovered claims.

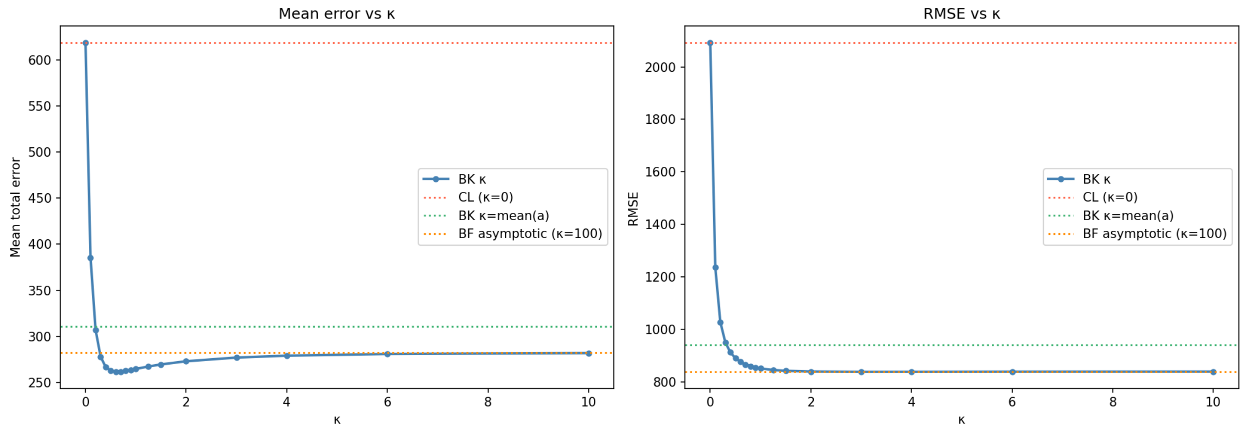

Figure 7: Mean prediction error (left) and RMSE (right) as a function of κ.

We notice that we get better results when we do this. Almost all parameters κ works well, and for lets say $\kappa = 1$ we can compare the distributions for the different methods in below Figure 8

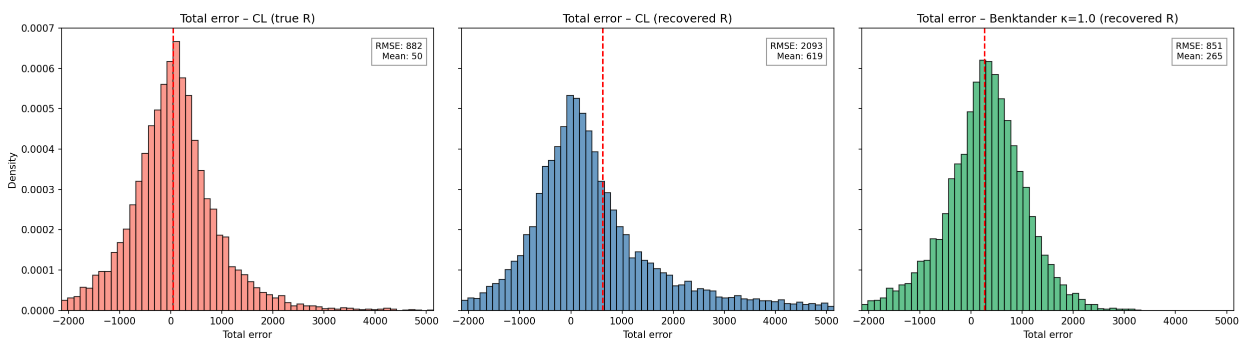

Figure 8: Comparison of the distribution of the prediction errors.

One remark is that the Benktander and Bornhuetter-Ferguson methods has the same quality in prediction result when applied to the processed claims, to that of chain-ladder on the recovered reportings, at least for $\eta = 1.1$. Both obtained a RMSE of around $1600 - 1900$ and a mean error of -360. Compare that to chain-ladder on processed claims (Table 2).

6.1 Known Labels

In this scenario we extend our assumptions and assume that the actuary obtain each label of a claim when processed. This means the actuary will partially know the reported claims at a given moment. At time t the actuary knows the label of $\sum_{i=1}^{t} P_i$ reported claims. If

$$T = \max\{i : \sum_{k=1}^{i} R_k \leq \sum_{k=1}^{t} P_k\},$$

then we know $R_{i,j}$ for $i + j \leq T$ and in addition the lower bound for $R_{i,j}$ when $i + j = T + 1$ can be refined. To recreate the unobserved reported claims the procedure is similar. We set up the minimization problem just as in (15)

$$\min_{\mathbf{r}\geq 0} \|C\,\mathbf{r}-\mathbf{d}\|^2\,, \quad C=\begin{bmatrix} M_r \\ \sqrt{\gamma}\,A \end{bmatrix}, \quad \mathbf{d}=\begin{bmatrix} M_p\,\mathbf{p} \\ \sqrt{\gamma}\,\mathbf{s} \end{bmatrix}.$$

However we do not only require $r \geq 0$, we also require that $r_{i,j} = R_{i,j}$ for $i+j \leq T$ and $r_{i,j} \geq q_{i,j}$ for $i+j = T+1$, where $0 \leq q_{i,j} \leq R_{i,j}$ is the number of claims out of the $R_{i,j}$ reported that has been processed up until time t. $r_{i,j}$ is short for the entry in the vector r that corresponds to $r_{i,j}$.

We study the affect of this assumption on reserving in a simulation. J_p is set as our upper bound, $\hat{J}$, $\gamma = 100$ is the penalty parameter used, $\tilde{J} = 3$ is our estimation of J as it is trivial to estimate J in this case, the capacity ratio is set to $\eta = 1.1$ and we simulate $10{,}000$ independent rounds. Just as in section 6 we study the quality of the prediction of total reported claims dynamically by varying κ.

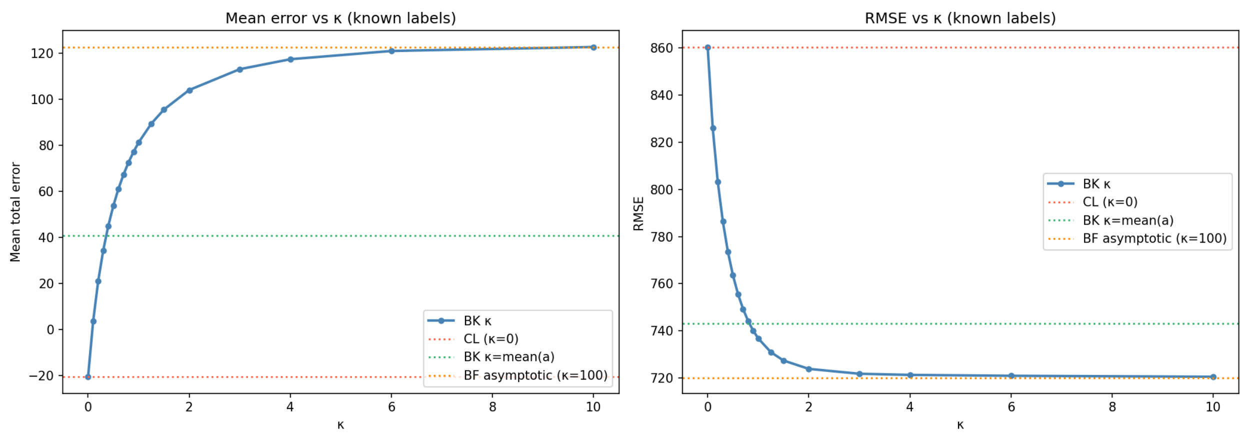

Figure 9: Plots of mean prediction error (left) and RMSE (right) as a function of κ.

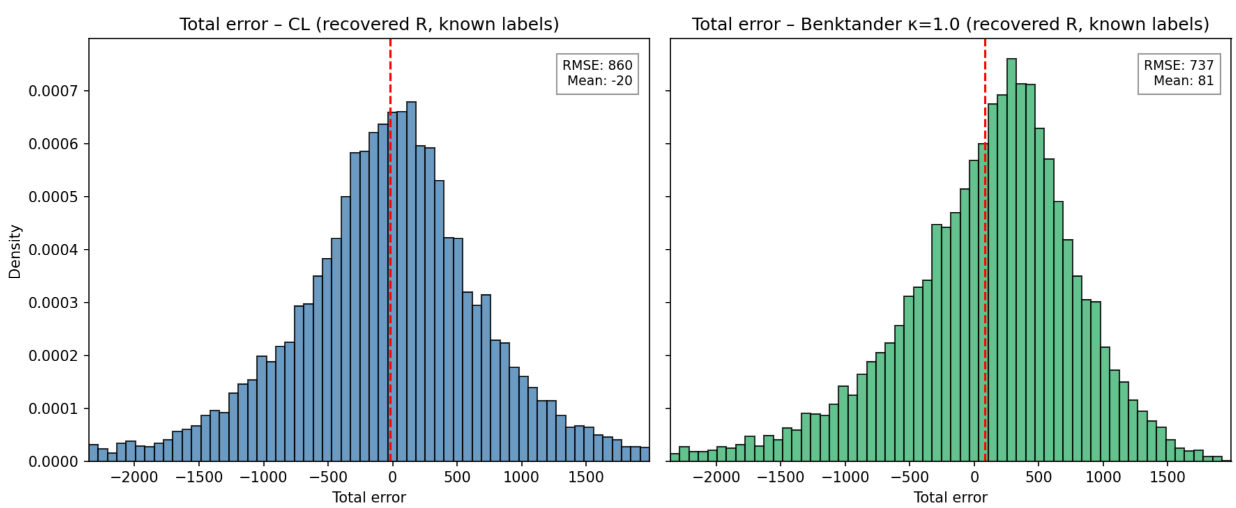

Figure 10: Comparison of the distribution of prediction errors.

Interestingly we see the opposite behavior for the mean prediction error in Figure 9 compared to the mean prediction error in Figure 7. Both however show a reduced RMSE when increasing κ. In addition the RMSE in this simulation is strikingly similar to that of the RMSE of the chain-ladder prediction in that of the other simulations, see Table 2 for example.

Table 5: RMSE - Recovery Methods

η	Method	$\gamma = 0$	$\gamma = 1.0$	$\gamma = 10$	$\gamma = 100$	CL(R)
1.1	CL($\hat{R}, J$)	6957	2487	2042	1994	902
	CL($\hat{R}, J_P$)	8385	79957	78805	79860	
	CL($\hat{R}, J_{P,\min}$)	8321	3361717*	70344	2476090*	
	CL($\hat{R}, J$) known labels	2819	852	859	860	
1.2	CL($\hat{R}, J$)	4417	1467	1199	1170	902
	CL($\hat{R}, J_P$)	4940	8938	9511	9501	
	CL($\hat{R}, J_{P,\min}$)	4924	8177	29847	37811	
	CL($\hat{R}, J$) known labels	1399	854	856	856	
1.4	CL($\hat{R}, J$)	1316	964	924	919	902
	CL($\hat{R}, J_P$)	2410	1077	1027	1020	
	CL($\hat{R}, J_{P,\min}$)	1738	1120	934	927	
	CL($\hat{R}, J$) known labels	1010	881	880	881	

Table 5 shows the RMSE for different configurations of capacity ratio and penalty parameter in recovery. It includes both cases where the label becomes known and not known for comparison. In the second case, the table has been completed with values from Table 2 and 4. We notice that the use of a non-zero penalty parameter, γ, works well for all tested values of η in the case where the label becomes known. If we suppose that the backlog is large enough such that no reported claims are processed the same calendar period they arrived for the last 4 calendar periods, then $b_t = b_{t-1} = b_{t-2} = b_{t-3} = 0$, and $r_{t,0}, r_{t-1,1}, r_{t-2,2}, r_{t-3,3}$ does not interact with the first term of Eq. (13)

$$\sum_{i=1}^{t}\sum_{j=0}^{t-i}\left(a_{i+j}\mathbb{1}_{\{j>0\}}\sum_{k=0}^{j-1}(r_{i,k} - P_{i,k}) + b_{i+j}r_{i,j} - P_{i,j}\right)^2 + \gamma\sum_{s=J}^{t}\left(\sum_{j=0}^{J} r_{s-j,j} - R_s\right)^2,$$

and will thus remain unchanged from the initialization of 0 if $\gamma = 0$, due to the gradient for those variables are 0 and they will never enter the free set $\mathcal{F}$ in the BVLS algorithm. This will result in an underestimate of $\sum_{j=0}^{3} R_{t,j}$, namely 0, when applying chain-ladder to the recovered claims.

7 Examples

7.1 Recovery

Below is an example of how the recovery of reported claims can look like for different methods we have used. We first began by simulating the true reportings and by applying the capacity ratio $\eta = 1.2$ we also simulated the FCFS queue to get the processed claims. The claims data for the latest 17 periods out of 250 is shown in Table 6.

True reported $R_{i,j}$			
120	116	32	0
2739	136	13	0
72	585	164	0
119	0	44	499
417	365	0	138
202	583	0	304
161	27	279	0
97	1086	64	16
833	44	55	0
266	122	75	0
1316	399	100	0
168	2	16	0
735	354	43	41
872	1024	514	34
430	143	177	
1887	13		
546			

Processed $P_{i,j}$					
120	30	37	21	48	12
719	689	695	734	51	0
0	0	44	725	52	0
0	0	119	218	325	0
0	293	489	138	0	0
0	383	402	152	138	14
58	130	130	141	8	0
97	515	525	117	9	0
403	396	105	28	0	0
0	319	121	23	0	0
637	944	234	0	0	0
98	72	16	0	0	0
735	344	40	28	17	
840	830	392	246		
330	143	73			
637	864				
0					

Table 6: True reported claims R (left) and processed claims P (right).

We notice that there is some backlog that has distorted the reported claims when we look at the processed ones. When an upper bound is chosen for J, the amount of non-zero columns is what we use. In Table 7 we see what the recovered reportings look like when using different estimates of J. If we estimate J by 3 the dimension of the output aligns with what we have simulated. If J_P is our estimate we obtain recovered reportings with at most as many columns as the processed claim data. Also seen in Figure 6, the usage of $J_{P,min}$ as our estimate often result in an underestimation of J and we get fewer columns when recovering. Lastly if we do not implement a penalty parameter for the diagonal sum, the entries $r_{17,0}, r_{16,1}, ..., r_{17-\tilde{J},\tilde{J}}$ on last diagonal is often 0 if the backlog is large enough, and if we have $\gamma \neq 0$, all entries obtained on the last diagonal are often of similar value, summing up to R_{17}, regardless of the method we choose to recover with.

$\hat{R}_{i,j}(J)$			
118	133	2	18
2723	143	96	0
23	440	319	0
163	0	0	466
263	416	219	0
196	474	101	303
85	84	288	0
78	1092	72	14
820	56	46	5
246	139	75	0
1310	415	87	1
147	19	17	1
731	356	43	41
867	1030	512	193
424	151	193	
1882	193		
193			

$\hat{R}_{i,j}(J_P)$				
120	119	0	37	0
2723	120	122	0	0
0	73	763	0	0
0	136	140	182	203
336	395	154	35	0
298	255	234	302	0
175	12	281	0	0
95	1084	66	17	0
834	44	54	0	0
242	161	57	3	0
1328	392	95	0	0
152	18	16	0	0
735	355	40	43	0
871	1026	514	0	
424	152	0		
1885	0			
0				

$\hat{R}_{i,j}(J_{P,\min})$		
118	133	16
2723	120	122
0	73	763
0	136	417
336	378	206
272	333	364
175	12	281
95	1084	80
834	44	54
246	153	64
1328	392	95
152	18	16
735	352	60
871	1026	514
424	152	0
1885	0	
0		

Table 7: Estimated $\hat{R}_{i,j}$ under $(J, \gamma = 100)$, $(J_P, \gamma = 0)$, and $(J_{P,\min}, \gamma = 0)$, left to right. Note the values on the last diagonal.

In the case that we obtain the label of every reported claim we process and adapt the minimization to what we know at the evaluation time, we see that the estimates for the true reportings gets better, for obvious reasons. However if there is enough backlog it can happen that all values on the latest diagonal will be of roughly the same value. See Table 8.

Known $R_{i,j}$			
120	116	32	0
2739	136	13	0
72	585	164	0
119	0	44	499
417	365	0	138
202	583	0	304
161	27	279	0
97	1086	64	16
833	44	55	0
266	122	75	0
1316	399	100	0
168	2	16	0
735	354	43	32
872	1024	412	0
430	116	0	
1501	0		
0			

Estimated $\hat{R}_{i,j}$			
120	116	32	0
2739	136	13	0
72	585	164	0
119	0	44	499
417	365	0	138
202	583	0	304
161	27	279	0
97	1086	64	16
833	44	55	0
266	122	75	0
1316	399	100	0
168	2	16	0
735	354	43	40
872	1024	515	193
430	144	192	
1885	193		
193			

Table 8: Known reportings $R_{i,j}$ at evaluation time (left) and estimated $\hat{R}_{i,j}$ ($\gamma = 100$) with the known reportings as lower bounds (right). Note the values on the last diagonal.

7.2 Recovery Failure

Sometimes especially for low capacity ratios the estimated reportings $R_{i,j}$ do not look similar to that of the true ones. To understand how the recovery fails in extreme cases is not trivial, neither is it trivial if the dimension of the problem is reduced, at least according to the author. Although it is interesting to see what the recovering looks like in extreme cases.

Table 9: Small example of how the estimated $R_{i,j}$ look like in extreme cases.

(a) True R

41	721	538	0
338	34	749	0
451	1707	342	0
220	450	525	3
45	123	397	0
27	654	4	0
1669	418	717	693
129	331	388	0
292	871	0	0
63	0	0	0

(b) Recovered $\hat{R}$

0	0	675	1267
0	579	599	0
579	811	838	677
0	0	0	2385
0	0	142	489
0	98	0	416
98	0	270	504
61	296	504	0
358	504	0	0
504	0	0	0

(c) Processed P

0	0	0	30	557	601	112	0
0	0	0	257	109	263	291	201
0	0	0	363	652	713	587	185
0	0	0	73	96	297	729	0
0	0	0	0	15	154	0	0
0	0	0	0	30	0	0	0
0	0	0	2	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

(d) a, b

a_i	b_i
0.3057	0
0.3129	0
0.3070	0
0.3137	0
0.2165	0
0.2283	0
0.2503	0
0.1828	0
0.2012	0
0.1927	0

Table 9 has been simulated with capacity ratio $\eta = 1.1$, penalty term $\gamma = 100$, we have estimated J by 3 and have 10 occurrence periods of interest. A clear warning sign that the recovery will not resemble the true reported claims is that of really high backlog. However it is not clear where the line is drawn.

8 Summary & Discussion

In this study we have analyzed different ways of estimating the true underlying reported claims that an actuary would be unaware of when reserving, especially when the insurance company has a claims processing unit that processes claims in a rate that is similar to that of incoming claims. We have studied the effect of different estimates of the non-deterministic development period for which no more claims are reported after, when recreating the unknown reported claims, in the sense of reserving with those estimates. The result was that the estimates we tried that were not equal to that of J we had chosen to simulate with did not yield satisfying results in terms of RMSE. Also when using the same J when recovering as when simulating had some weakness. Namely that the chain-ladder estimate often over-estimates $\sum_{i=1}^{17}\sum_{j=0}^{J} R_{i,j}$. The reason for this is unclear, but a large backlog is a good indication, however where we draw the line is still unclear and not trivial. We studied the effect of the diagonal sum penalty parameter γ, which was used to enforce $\sum_{j=0}^{J} R_{t-j,j} = R_t$. In conclusion, it helped with the reserving when it was non-zero in the case of estimating J correctly, but failed otherwise when sometimes giving unreasonable estimates, see Tables 3 and 4. In addition, we saw the effects of using more observed data than that of the periods we were interested in recovering. The results indicated that they were favorable in the sense of reducing bias.

We have also tried using different reserving methods. All of them can be categorized as Benktander Golden stairs, but with a twist. z in (16) is considered as a function of implicitly the development period, or the development factors to be more exact. We extended z to be a function of the backlog. The results were that it was favorable to rely more on the Bornhuetter-Ferguson estimate of ultimate claims when the backlog was high, compared to solely relying on the chain-ladder estimate. We noticed that it worked quite well and was comparable to chain-ladder on the simulated reported claims in terms of root mean square error. When we assumed the actuary obtained each occurrence period, development period and the timestamp of each claim once it was processed we saw that the reserving for all methods worked well according to our standard method of reserving with CL on the simulated reported claims, and better when only taking RMSE into account. In contrast to when we did not assume that the label became known when processed, we saw that the mean of the prediction error increased when κ increased. However RMSE as a function of κ did decrease.

For future work on this problem, one could try different parameter configurations on the random variables $R_{i,j}$ were simulated from, or use another distribution than the negative binomial. One could also look deeper into when and why the estimated reported claims differ substantially from the true ones. Or see what happens when the capacity is more dynamic or random.

References

[1] Filip Lindskog and Mario V. Wüthrich. Eliciting claims development patterns and costs hidden in backlogs. *European Actuarial Journal*, 15:667–705, 2025.

[2] Thomas Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin: The Journal of the IAA*, 23(2):213–225, 1993.

[3] Walther Neuhaus. The benktander golden stairs and other parameter-free credibility methods in loss reserving. *Scandinavian Actuarial Journal*, 2026(4):392–411, 2026.

[4] Philip B Stark and Robert L Parker. Bounded-variable least-squares: an algorithm and applications. *Computational Statistics*, 10:129–129, 1995.