

Claims processing and costs under capacity constraints

Filip Lindskog and Mario V. Wüthrich

September 17, 2024

Abstract

Random delays between the occurrence of accident events and the corresponding reporting times of insurance claims is a standard feature of insurance data. The time lag between the reporting and the processing of a claim depends on whether the claim can be processed without delay as it arrives or whether it remains unprocessed for some time because of temporarily insufficient processing capacity that is shared between all incoming claims. We aim to explain and analyze the nature of processing delays and build-up of backlogs. We show how to select processing capacity optimally in order to minimize claims costs, taking delay-adjusted costs and fixed costs for claims settlement capacity into account. Theoretical results are combined with a large-scale numerical study that demonstrates practical usefulness of our proposal.

Keywords: claims processing, backlog, capacity constraints

JEL codes: G22; G31

1 Introduction

Accident events give insurance policyholders the right to financial compensation. In most cases, the effective costs are not immediately known to the insurance company. Reporting delay is a standard feature of insurance data, and once a claim is reported, it is not necessarily settled quickly since claims may take time to process. With an unlimited processing capacity, claims can be processed fast(er), however, economic reasons allow insurance companies to only allocate a limited capacity to the claims settlement unit (process). Naturally, this capacity should be bigger than the average claims volume, otherwise there will be growing an infinitely large backlog of unprocessed claims. This is a consideration on average, which is essentially distorted by the fact that claims occurrence and reporting can cluster, i.e., there may be peaks of claims reportings, but also quiet periods where reportings are below average. The question we study is how can the capacity be set optimally so that backlogs of processing are not too large, and at the same time periods of low reportings do still not lead to very quiet times in claims handling units. The latter implies high fixed costs, as an

inactive claims handling unit still needs to be compensated. Generally too high backlogs also lead to additional costs, it is verified that late claims settlements typically increase the claims costs. Thus, there is a trade-off in costs between low and high claims processing capacity, and our aim is to study an optimal balance between the two.

The study of systems where constraints on processing capacity induces processing delays and dependence is at the heart of queueing theory. If we would be content with studying the system with incoming reported claims and outgoing processed claims without labeling the incoming reported claims, then our study could essentially be reduced to an application of standard queueing theory. However, insurance applications require the input to be labeled in the sense that incoming reported claims belong to different contract groups and we need to keep track of the evolution of processed claims for each such group that shares the processing capacity. This topic seems new to the actuarial literature because we did not find any literature on this topic. It has come to our attention because we have been approached by an insurer facing a backlog of unprocessed claims that needed to be worked off in a cost efficient way. This question is related to queueing theory from where we borrow some mathematical results. Nevertheless, many parts significantly differ from queueing theory, mathematically as well as from an interpretation and terminology point of view.

Optimal capacity sizing for systems where users share the system's capacity has been studied in the operations research literature. Optimality may be considered in terms of stability of the system or in terms of maximization of profits generated by the system. The study by Maglaras and Zeevi [13] provides one example from this area of research.

In the actuarial literature there does not seem to exist works that study problems close to the one we consider. However, economic consequences of delays in claims settlement has indeed been studied. Boogaert and Haezendonck [4] consider claims arriving according to a homogeneous Poisson process. To the sequence of claims an i.i.d. sequence of triplets (X_n, H_n, V_n) is added, where X_n is the claim size, H_n the handling delay, and V_n the payment delay. By considering an economic environment with time-varying inflation and interest rate, the present value of insurance liabilities is affected by possible dependence between the elements of the triplet (X_n, H_n, V_n) , such as positive dependence between claim size and handling delay. Huynh et al. [9] consider reported claims according to a compound Poisson model where incoming claims are either processed (and paid) immediately or investigated by the claim handler. Incoming claims are handled independently with equal probability of being investigated. An investigation causes a delay in the processing of the claim and also a claim cost whose distribution differs from that of claims that do not undergo investigation. The resulting surplus process can be seen as the output from a queueing system. The effects on the ruin probability of investigating claims, through delayed processing and modified claim cost, are investigated. Related studies of effects of delays in claims settling on ruin probabilities are Waters and Papa- triandafylou [16] and Albrecher et al. [1]; see also references therein. Although there are studies on effects of processing and payment delays on liability values

or ruin probabilities, we have not found literature that analyzes effects of capacity constraints on processing delays and delay-adjusted claims costs for contract groups that share processing capacity.

The paper is organized as follows. Section 2 defines the key variables for our study, motivates their relationships and explains the queueing system that arises. Section 3 discusses costs for processing capacity, delay-adjusted claims costs, and introduces joint models for combined delay-adjusted claims costs and settlement costs that form the basis of the minimization problems we study. Section 4 introduces the procedure for how claims are processed, taking into account that at each time different contract groups share the current processing capacity. Section 5 derives expressions for unconditional and conditional expectations of the number of claims in the backlog for different contract groups. These expectations are key ingredients in the cost minimization problems. Section 6 introduces the stochastic model that allows us to consider a realistic large-scale application of the framework and results presented earlier. Section 7 explains in detail how we approximate terms appearing in the expressions for the backlog expectations by recurrent neural networks. Section 8 solves the cost minimization problems numerically. Finally, we summarize in Section 9.

2 Reported, processed, and backlog claims

2.1 Definitions and assumptions

We index the number of reported (R), processed (P), and backlog (B) claims by occurrence period and development period. Let $\{i_0, i_0 + 1, \dots, 1, 2, \dots\}$ be the index set for occurrence periods and let $\{0, 1, \dots\}$ be the index set for development periods. Periods may refer to months, quarters, years, etc. The general index sets allow us to define and study the numbers of reported, processed and backlog claims as stochastic processes.

Occurrence period i starts at (calendar) time $i-1$ and ends at (calendar) time i . Let $R_{i,j}$ denote the number of reported claims due to accident events during occurrence period i that are reported during development period j . Hence, these reportings occur between time $i-1+j$ and time $i+j$. We assume the existence of a non-random integer J such that $R_{i,j} = 0$ for any $j > J$. Hence, we assume a maximal reporting delay of $J+1$. Let $P_{i,j}$ denote the number of processed claims due to events during occurrence period i that are processed during development period j . A claim cannot be processed before it has been reported. $R_{i,j}$ and $P_{i,j}$ are observable at time $i+j$. Let $B_{i,j}$ denote the number of backlog claims due to events during occurrence period i that are already reported but in the backlog and, therefore, have not yet been processed by development period j . Set $B_{i,0} \equiv 0$, meaning that each new occurrence period starts without any backlog. $B_{i,j}$ is observable at time $i+j$ (and may be inferred at time $i+j-1$ depending on what other variables are observed, see (3) below). Let $C_t > 0$ denote the number of claims that the insurer has the capacity to process during time period t , i.e., the time period between time $t-1$ and t . In

general, C_t is a random variable.

The total number of reported claims during time period t is

$$R_t := \sum_{i=t-J}^t R_{i,t-i} = \sum_{j=0}^J R_{t-j,j}.$$

The total number of processed and backlog claims, respectively, during time period t are

$$P_t := \sum_{i=i_0}^t P_{i,t-i} = \sum_{j=0}^{t-i_0} P_{t-j,j},$$

$$B_t := \sum_{i=i_0}^t B_{i,t-i} = \sum_{j=0}^{t-i_0} B_{t-j,j}.$$

There is sufficient capacity during time period t to process all not-yet-processed claims reported during period t or earlier if the event SC_t , given by

$$SC_t := \{B_t + R_t \leq C_t\},$$

is true. The inequality means that there is sufficient capacity to process both the backlog B_t and the newly reported claims R_t . If there is not sufficient capacity, SC_t^c , some backlog is forwarded to the next time period.

We emphasize some general properties for reported, processed, and backlog claims. Each reported claim will at some point be processed, possibly by temporarily contributing to the backlog. A processed claim remains processed. A claim in the backlog either remains temporarily in the backlog or transitions into a processed claim (terminal state). In mathematical terms, any procedure for processing claims should satisfy the following axioms (1)-(4):

$$R_{i,j} \geq 0, B_{i,j} \geq 0, B_{i,0} = 0, P_{i,j} \geq 0, \text{ for all } i, j, \quad (1)$$

$$\sum_{j=0}^{\infty} R_{i,j} = \sum_{j=0}^{\infty} P_{i,j}, \text{ for all } i, \quad (2)$$

$$B_{i,j+1} = B_{i,j} + R_{i,j} - P_{i,j}, \text{ for all } i, j. \quad (3)$$

An immediate consequence of (1) and (3) is

$$\sum_{j=0}^k R_{i,j} = B_{i,k+1} + \sum_{j=0}^k P_{i,j}, \text{ for all } i, k.$$

The total number of processed claims during time period t is

$$P_t = \min(B_t + R_t, C_t), \quad (4)$$

i.e., the sum of the numbers of backlog and reported claims if the sum does not exceed the capacity C_t of period t , otherwise C_t . From (3) it follows that $B_{t+1} = B_t + R_t - P_t$ which together with (4) give

$$B_{t+1} = \max(B_t + R_t - C_t, 0). \quad (5)$$

The recursion (5) for the total size of the backlog is an example of the Lindley recursion that is well studied in queueing theory, see Example I.5.7 and Chapter III.6 in [2]. If $(R_t)_{t=0}^\infty$ and $(C_t)_{t=0}^\infty$ are i.i.d. sequences, the Lindley process $(B_t)_{t=0}^\infty$ is a Markov process given by

$$B_0 = b, \quad B_{t+1} = \max(B_t + R_t - C_t, 0), \quad t \geq 0,$$

which is the waiting time process for a GI/G/1 queue. If $\mathbb{E}[R_t] < \mathbb{E}[C_t]$ and $\mathbb{E}[(R_t - C_t)^2] < \infty$, there is a stationary distribution with finite mean to which B_t converges in distribution as $t \rightarrow \infty$.

Assumption 2.1. The stochastic system $\{(B_{i,j}, R_{i,j}, P_{i,j}, C_{i+j}) : i \geq i_0, j \geq 0\}$ satisfies (1), (2), (3) and (4).

To understand the stochastic system, we specify σ -algebras that play a natural role in conditional probabilities and expectations that will appear. Let

$$\begin{aligned} \mathcal{E}_t &:= \sigma(R_s, B_s : s \leq t), \\ \mathcal{F}_t &:= \sigma(R_s, B_s, P_s : s \leq t), \\ \mathcal{G}_t &:= \sigma(R_{i,j}, B_{i,j} : i+j \leq t, j \geq 0), \\ \mathcal{H}_t &:= \sigma(R_{i,j}, B_{i,j}, P_{i,j} : i+j \leq t, j \geq 0). \end{aligned}$$

By construction, the σ -algebras obviously satisfy $\mathcal{E}_t \subset \mathcal{F}_t \subset \mathcal{H}_t$ and $\mathcal{E}_t \subset \mathcal{G}_t \subset \mathcal{H}_t$. From (3), it follows that $\mathcal{F}_t \subset \mathcal{E}_{t+1} \subset \mathcal{F}_{t+1}$ and $\mathcal{H}_t \subset \mathcal{G}_{t+1} \subset \mathcal{H}_{t+1}$. Note that B_{t+1} is \mathcal{F}_t -measurable but in general not \mathcal{E}_t -measurable. Similarly, $B_{i,j+1}$ is \mathcal{H}_{i+j} -measurable but in general not \mathcal{G}_{i+j} -measurable. From (3) and (4), it follows that $SC_t \in \mathcal{F}_t$. However, we emphasize that from Assumption 2.1 alone it does not follow that C_t is measurable w.r.t. any of the σ -algebras $\mathcal{E}_t, \mathcal{F}_t, \mathcal{G}_t, \mathcal{H}_t$. This is not surprising: if we are not observing the capacity C_t but only its effect on the number of backlog claims and processed claims, then the actual capacity may be larger than the capacity used to the process claims. However, if C_t is non-random, then (3) and (4) together imply that B_{t+1} is \mathcal{E}_t -measurable.

Remark 2.2. In queueing theory, the Lindley recursion (5) describes the waiting time B_t for the t th customer arriving to a single service station, with R_t the service time for the t th customer and C_t the time between the arrival of the t th customer and that of the $(t+1)$ th customer. If both (R_t) and (C_t) are i.i.d. sequences, the queueing system is denoted GI/G/1. The special case when both service times R_t and inter-arrival times C_t are exponentially distributed is denoted M/M/1. The special case when $C_t = c$ is constant is denoted D/G/1. We emphasize that the interpretation of the variables B_t, R_t and C_t in our setting is quite different although many results known for GI/G/1 queues are

also useful to understand the dynamics of the total number of backlog claims. The expectation $\mathbb{E}[B]$ of the stationary distribution of (B_t) can in general not be obtained explicitly. However, it can be approximated numerically; see, e.g., [10] and [11] for the analysis of the D/G/1 queue.

2.2 Stationary behavior

The recursion (5) for the total size of the backlog is an example of the Lindley recursion that is well studied in queueing theory, see Chapter III.6 [2]. As a direct consequence of the recursion (5),

$$B_{t+1} = \max \left(B_0 + \sum_{s=0}^t (R_s - C_s), \sum_{s=1}^t (R_s - C_s), \dots, R_t - C_t, 0 \right), \quad t \geq 0.$$

The asymptotic behavior of B_{t+1} is well understood when $(R_s - C_s)_{s=0}^\infty$ is an i.i.d. sequence with $\mathbb{E}[R_s] < \mathbb{E}[C_s]$. If $(C_t)_{t=0}^\infty$ is an i.i.d. sequence, if all $R_{i,j}$ are independent, and if $R_{i',j} \stackrel{d}{=} R_{i,j}$ for all j , then $(R_s - C_s)_{s=0}^\infty$ is an i.i.d. sequence if $i_0 \leq -J$. We write $\mathbb{E}[R]$ for the common expected value for any element of the i.i.d. sequence $(R_s)_{s=0}^\infty$. Note that in this case, $\mathbb{E}[R] = \sum_{j=0}^J \mathbb{E}[R_{i,j}]$ is simply the expected total number of reported claims for any occurrence period i . By Corollary III.6.5 in [2], we have convergence in distribution

$$B_t \xrightarrow{d} \max_{0 \leq s < \infty} \left(0, \sum_{r=0}^s (R_r - C_r) \right) \quad \text{as } t \rightarrow \infty.$$

From $\sum_{r=0}^s (R_r - C_r) \rightarrow -\infty$, a.s., as $s \rightarrow \infty$, it follows that the limit variable is well defined. Dropping the subscript, we denote by B a random variable whose distribution is the stationary distribution of $(B_t)_{t=0}^\infty$. By Proposition VIII.4.5 [2],

$$\mathbb{E}[B] = \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}[\max(S_k, 0)], \quad S_k := \sum_{s=0}^{k-1} (R_s - C_s).$$

Unfortunately, it is rarely possible to compute $\mathbb{E}[B]$ explicitly and numerical evaluation is non-trivial in general; see, e.g., [15] for an approach to computing stationary probabilities for integer-valued B .

Upper bounds for $\mathbb{E}[B]$ are studied in [6] by considering a scaled version of the Lindley recursion in the setting of GI/G/1 queues. Since (R_t) and (C_t) are i.i.d. sequences in the GI/G/1 queueing setting, we write R and C for an arbitrary element of these sequences. The recursion (5) can be written

$$\tilde{B}_{t+1} = \max(\tilde{B}_t + \tilde{R}_t - \tilde{C}_t, 0),$$

where $\tilde{B}_t := B_t/\mathbb{E}[C]$, and similarly for \tilde{R}_t and \tilde{C}_t . By construction $\mathbb{E}[\tilde{C}_t] = 1$, and $\mathbb{E}[\tilde{B}_t] = \mathbb{E}[B]/\mathbb{E}[C]$ and $\mathbb{E}[\tilde{R}_t] = \mathbb{E}[R]/\mathbb{E}[C] =: \rho$. The parameter ρ

is called the traffic intensity in queueing theory. The so-called heavy-traffic approximation of $\mathbb{E}[B]$ (obtained by considering the behavior as $\rho \rightarrow 1$) is

$$\frac{\mathbb{E}[B]}{\mathbb{E}[C]} \approx \frac{\rho^2}{2(1-\rho)} \left(\frac{\text{Var}[R]}{\mathbb{E}[R]^2} + \frac{\text{Var}[C]}{\mathbb{E}[C]^2} \right), \quad (6)$$

this is equivalent to expression (2.9) in [6]. In the D/G/1 setting $\text{Var}[C] = 0$, since $C = c$ is constant, and (6) takes the form

$$\mathbb{E}[B] \approx \frac{\mathbb{E}[R]}{c - \mathbb{E}[R]} \frac{\text{Var}[R]}{2\mathbb{E}[R]}. \quad (7)$$

It can be shown ((2.6) and (2.7) in [6]) that the right-hand side in (7) coincides with the upper bounds for $\mathbb{E}[B]$ obtained by Kingman in [12] and by Daley in [7]. In the M/G/1 setting, where C is exponentially distributed with mean c , $\text{Var}[C]/\mathbb{E}[C]^2 = 1$ and the heavy-traffic approximation for $\mathbb{E}[B]$ in (6) equals

$$\frac{c\rho^2}{2(1-\rho)} \left(\frac{\text{Var}[R]}{\mathbb{E}[R]^2} + 1 \right) = \frac{\mathbb{E}[R]}{c - \mathbb{E}[R]} \frac{\mathbb{E}[R^2]}{2\mathbb{E}[R]},$$

which, in fact, is an exact expression for $\mathbb{E}[B]$ referred to as the Pollaczek-Khintchine formula; see VIII.(5.6) in [2].

Remark 2.3. In the D/G/1 setting, an upper bound for $\mathbb{E}[B]$ is given by the right-hand side in (7). If we write $c = \eta\mathbb{E}[R]$ for some $\eta > 1$, then

$$\mathbb{E}[B] \leq \frac{1}{\eta - 1} \frac{\text{Var}[R]}{2\mathbb{E}[R]}.$$

If $\mathbb{E}[R]$ is fairly large (which is the case for realistic insurance applications), then, unless $\eta \approx 1$, $\text{Var}(R)$ needs to be substantially larger than $\mathbb{E}[R]$ in order for $\mathbb{E}[B]$ and $\mathbb{E}[R]$ to be of similar size. In particular, if R is Poisson distributed, then $\text{Var}[R] = \mathbb{E}[R]$, and no substantial backlog will appear unless $\eta \approx 1$, corresponding to a close to non-stationary system. In our examples below, we consider a negative binomial distribution for the number of reported claims R .

3 Cost implications of capacity constraints

Capacity constraints are mainly due to limited financial resources, and considerations of how much of these financial resources should be allocated to the claims settlement unit. Claims settlement costs are belonging to the unallocated loss adjustment expenses (ULAE) meaning that these costs are not specific to an individual claim, but they are rather overhead costs that are necessary to run the claims handling unit. ULAE then need to be allocated to occurrence periods (or individual insurance contracts) in order to be able to perform a profit analysis for occurrence periods (or individual insurance contracts); see Buchwalder et al. [5] for a method supporting the chain-ladder claims reserving method. In

most cases, the allocation is chosen to be proportional either to claims costs, claims counts or a linear combination of the two.

We will perform an analysis of expected claims costs where we consider unconditional expectations of the numbers of reported, processed, and backlog claims. This means that we consider i.i.d. sequences $(R_t)_{t=0}^{\infty}$ and $(C_t)_{t=0}^{\infty}$ under the stochastic system in Assumption 2.1 in its stationary state. In particular, the backlog process $(B_t)_{t=0}^{\infty}$ is given by (5) with initial state B_0 drawn from the stationary distribution, see Section 2.2. We write $\mathbb{E}[R]$ and $\mathbb{E}[B]$ for the expected number of reported claims and size of the backlog, respectively, in any given period for the stationary system. We write $\mathbb{E}[C]$ for the expected maximal number of claims that can be processed in any given period. Recall that $\mathbb{E}[C] > \mathbb{E}[R]$ is assumed since we are considering the system in its stationary state. We further assume that claims costs for individual claims form an i.i.d. sequence independent of the stochastic system in Assumption 2.1.

Claims settlement costs. Making claims processing capacity available generates ULAE and these expenses need to be allocated to occurrence periods to have an integrated cost view. Suppose that any claims occurrence period τ uses the processing capacity during periods $\tau, \tau + 1, \dots, \tau + J_P - 1$, where $J_P \geq 1$. The cost allocated to occurrence period τ is the total expected cost for these processing capacities multiplied by the fraction of expected number of processed claims for occurrence period τ during these periods divided by the expected number of processed claims for all occurrence periods that share the capacity during these periods. Let us formalize this. The set of index pairs (i, j) for the number of processed claims $P_{i,j}$ during the periods $\tau, \tau + 1, \dots, \tau + J_P - 1$ is

$$\mathcal{I} := \{(i, j) : \tau \leq i + j \leq \tau + J_P - 1, 0 \leq j \leq J_P - 1\},$$

and due to stationarity of the numbers of processed claims ($\mathbb{E}[P_{i,j}] = \mathbb{E}[P_{i',j}]$) we may sum over rows instead of over diagonals

$$\begin{aligned} \sum_{(i,j) \in \mathcal{I}} \mathbb{E}[P_{i,j}] &= \sum_{k=0}^{J_P-1} \sum_{i=\tau+k-J_P+1}^{\tau+k} \mathbb{E}[P_{i,\tau+k-i}] \\ &= \sum_{i=\tau}^{\tau+J_P-1} \sum_{j=0}^{J_P-1} \mathbb{E}[P_{i,j}] = J_P \sum_{j=0}^{J_P-1} \mathbb{E}[P_{\tau,j}]. \end{aligned}$$

We see that to any occurrence period τ we should allocate a fraction $1/J_P$ of the cost for processing capacity during J_P periods. The integer J_P cancels out when multiplying the two numbers and we conclude that the capacity cost allocated to any occurrence period equals the full cost for processing capacity during a single period. We emphasize that this is the stationary case.

Linearly delay-adjusted claims costs. The total expected ground-up claims costs for any occurrence period i in stationarity is

$$\kappa_g \sum_{j=0}^J \mathbb{E}[R_{i,j}] = \kappa_g \sum_{j=0}^J \mathbb{E}[R_{t-j,j}] = \kappa_g \mathbb{E}[R],$$

where $\kappa_g > 0$ denotes the expected claims cost for an individual claim if paid without any delay. We assume that delayed processing generally makes claims more expensive. In the most simple linear model (ℓ) we assume an additional constant $\kappa_b > 0$ that originates from late processing. The expected total claims costs of occurrence period i are then in this linear cost model given by

$$\kappa_g \mathbb{E}[R] + \sum_{j \geq 0} \kappa_b \mathbb{E}[B_{i,j}] = \kappa_g \mathbb{E}[R] + \sum_{j \geq 0} \kappa_b \mathbb{E}[B_{t-j,j}] = \kappa_g \mathbb{E}[R] + \kappa_b \mathbb{E}[B],$$

where we used the assumption of the system in stationarity.

Non-linearly delay-adjusted claims costs. Alternatively, we could consider a claims cost model where we rather have the view of an inflation-adjusted cost (ι). In that case we consider an additional constant $\lambda_b > 1$ and set

$$\kappa_g \sum_{j \geq 0} \lambda_b^j \mathbb{E}[P_{i,j}] = \kappa_g \sum_{j \geq 0} \lambda_b^j \mathbb{E}[B_{i,j} + R_{i,j} - B_{i,j+1}],$$

with $R_{i,j} \equiv 0$ for $j > J$. This model considers processed claims and inflates ground-up costs $\kappa_g \lambda_b^j$ by its processing delay from the end of the occurrence period. Note that λ_b^j should be seen as a super-imposed delay inflation which is different from economic inflation. In fact, by assuming constant ground-up costs we implicitly assume that all costs have been adjusted for economic inflation so that they live on the same scale, and additional backlog costs are then concerned with super-imposed claims inflation and costs related to an increased expense due to late processing and settlements.

Combining delay-adjusted claims costs and settlement costs. Making claims processing capacity with expectation $\mathbb{E}[C]$ available generates ULAE. We have explained above that the ULAE allocated to any given occurrence period corresponds to the full single-period ULAE. For simplicity, we assume that ULAE have a fixed component that corresponds to a minimal expected capacity $\mathbb{E}[R]$ ensuring a stationary system, and for the excess expected capacity $\mathbb{E}[C] - \mathbb{E}[R]$ we consider a proportional cost $\kappa_c (\mathbb{E}[C] - \mathbb{E}[R])$ with $\kappa_c > 0$. Adding the expected costs for the excess capacity to the delay-adjusted expected claims costs gives the following expected costs for any occurrence period i :

$$\mu_i^{(\ell)} := \kappa_g \mathbb{E}[R] + \kappa_b \mathbb{E}[B] + \kappa_c (\mathbb{E}[C] - \mathbb{E}[R]), \quad (8)$$

for the linear cost model, and

$$\mu_i^{(\iota)} := \kappa_g \sum_{j \geq 0} \lambda_b^j \mathbb{E}[B_{i,j} + R_{i,j} - B_{i,j+1}] + \kappa_c (\mathbb{E}[C] - \mathbb{E}[R]), \quad (9)$$

for the model with non-linear delay-inflated costs. We emphasize that $\mu_i^{(\ell)}$ for the linear cost model does not depend on how processing capacity is shared between occurrence periods requiring processing capacity (the indexes i and j do not show up in the expression (8)). However, $\mu_i^{(\iota)}$ for the non-linear cost model does indeed depend on how processing capacity is shared. It is known that $\mathbb{E}[B]$

is a convex function of $\mathbb{E}[C]$, see, e.g., [8], and therefore $\mu_i^{(\ell)}$ is a convex function of $\mathbb{E}[C]$. The main question then is what is the optimal expected capacity to minimize the overall expected costs. In general, this expected cost minimization has to be performed numerically, and we return to this in Section 8.1.

Example 3.1. As shown in Section 2.2, a heavy-traffic approximation for $\mathbb{E}[B]$ may give an explicit expression for $\mu_i^{(\ell)}$ of the form

$$\kappa_g \mathbb{E}[R] + \kappa_b \mathbb{E}[B] + \kappa_c (\mathbb{E}[C] - \mathbb{E}[R]) = \kappa_g \mathbb{E}[R] + \kappa_b \frac{\alpha \mathbb{E}[R]}{c - \mathbb{E}[R]} + \kappa_c (c - \mathbb{E}[R]),$$

which can be minimized explicitly with minimizer $c = \mathbb{E}[R] + \sqrt{\alpha \mathbb{E}[R] \kappa_b / \kappa_c}$ and minimum $\kappa_g \mathbb{E}[R] + 2\sqrt{\alpha \kappa_b \kappa_c \mathbb{E}[R]}$. The D/G/1 queueing model setting corresponds to $\alpha = \text{Var}[R] / (2\mathbb{E}[R])$ which leads to the minimizer $c = \mathbb{E}[R] + \sqrt{\text{Var}[R] \kappa_b / (2\kappa_c)}$ and minimum $\kappa_g \mathbb{E}[R] + \sqrt{2\kappa_b \kappa_c \text{Var}[R]}$.

4 Sharing processing capacity

Many procedures can be considered describing how claims are processed and how the backlog for individual occurrence periods evolves over time. We focus on the simple procedure where claims are processed by first processing the backlog and then, if there is processing capacity left after the backlog has been processed, the newly reported claims are processed. However, we do this without assuming continuous-time monitoring of claim arrivals (reporting times). Let

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

denote the event that there is, at time t , sufficient capacity to process the backlog. Note that

$$\text{SC}_t^c \cap \text{SCB}_t = \{B_t \leq C_t < B_t + R_t\}$$

denotes the event that the capacity is insufficient to process all claims waiting to be processed, but sufficient to process the backlog.

The number of processed claims is the sum of the number of processed backlog claims and the number of processed newly reported claims

$$P_{i,t-i} = P_{i,t-i}^B + P_{i,t-i}^R. \quad (10)$$

We assume that, given $\mathcal{F}_t \vee \mathcal{G}_t := \sigma\{\mathcal{F}_t, \mathcal{G}_t\}$, claims in the backlog at the beginning of period t will be processed during period t independently with (conditional) probability

$$\mathbb{1}_{\text{SCB}_t} + \frac{C_t}{B_t} \mathbb{1}_{\text{SCB}_t^c}. \quad (11)$$

From (3) and (4) it follows that this expression for the conditional probability is \mathcal{F}_t -measurable. This is seen as follows. If $P_t \geq B_t$, then $\mathbb{1}_{\text{SCB}_t} = 1$ and

$C_t \mathbb{1}_{\text{SCB}_t^c} = 0$. If $P_t < B_t$, then $\mathbb{1}_{\text{SCB}_t} = 1$ and $C_t \mathbb{1}_{\text{SCB}_t^c} = P_t$. Hence, the conditional probability is fully determined by P_t and B_t being both \mathcal{F}_t -measurable.

The interpretation of the conditional probability (11) is straightforward: If there is sufficient capacity to process the backlog, then the probability is equal to one. Otherwise the probability equals the proportion of the capacity to the size of the backlog. Given $\mathcal{F}_t \vee \mathcal{G}_t$, we know the number $B_{i,t-i}$ of backlog claims for occurrence period i . Hence, it follows from the assumption of independently processing the backlog claims that the conditional distribution $\mathcal{L}(P_{i,t-i}^B \mid \mathcal{F}_t \vee \mathcal{G}_t)$ is a binomial distribution. Therefore,

$$\mathbb{E}[P_{i,t-i}^B \mid \mathcal{F}_t \vee \mathcal{G}_t] = B_{i,t-i} \left(\mathbb{1}_{\text{SCB}_t} + \frac{C_t}{B_t} \mathbb{1}_{\text{SCB}_t^c} \right). \quad (12)$$

We assume that, given $\mathcal{F}_t \vee \mathcal{G}_t$, claims reported during period t will be processed during period t independently with (conditional) probability

$$\mathbb{1}_{\text{SC}_t} + \frac{C_t - B_t}{R_t} \mathbb{1}_{\text{SC}_t^c \cap \text{SCB}_t}. \quad (13)$$

From (3) and (4) it follows that this expression for the conditional probability is \mathcal{F}_t -measurable. That is, if $P_t = B_t + R_t$, then $\mathbb{1}_{\text{SC}_t} = 1$. Otherwise, if $P_t < B_t + R_t$, then $\mathbb{1}_{\text{SC}_t} = 0$. If $P_t < B_t + R_t$ and $P_t \geq B_t$, then $P_t = C_t \mathbb{1}_{\text{SC}_t^c \cap \text{SCB}_t}$. Otherwise, if $P_t = B_t + R_t$ or $P_t < B_t$, then $\mathbb{1}_{\text{SC}_t^c \cap \text{SCB}_t} = 0$. Hence, the conditional probability is fully determined by P_t , B_t and R_t which are all \mathcal{F}_t -measurable.

The interpretation of the conditional probability (13) is as follows: If there is sufficient capacity to process first the backlog and then the newly reported claims, then the probability is equal to one. Otherwise the probability equals the proportion of the remaining capacity (after processing the backlog) to the number of reported claims. Similarly to above for the backlog claims, the newly reported claims are processed independently with the corresponding remaining capacity, giving another binomial distribution. Thus, given $\mathcal{F}_t \vee \mathcal{G}_t$, we know the number $R_{i,t-i}$ of reported claims for occurrence period i , the conditional distribution $\mathcal{L}(P_{i,t-i}^R \mid \mathcal{F}_t \vee \mathcal{G}_t)$ is a binomial distribution and

$$\mathbb{E}[P_{i,t-i}^R \mid \mathcal{F}_t \vee \mathcal{G}_t] = R_{i,t-i} \left(\mathbb{1}_{\text{SC}_t} + \frac{C_t - B_t}{R_t} \mathbb{1}_{\text{SC}_t^c \cap \text{SCB}_t} \right). \quad (14)$$

Equations (10), (12) and (14) together define the procedure for processing claims. By summing up the number of processed backlog claims and the number of processed newly reported claims we obtain the conditionally expected number of processed claims

$$\begin{aligned} \mathbb{E}[P_{i,t-i} \mid \mathcal{F}_t \vee \mathcal{G}_t] &= B_{i,t-i} \left(\mathbb{1}_{\text{SCB}_t} + \frac{C_t}{B_t} \mathbb{1}_{\text{SCB}_t^c} \right) \\ &\quad + R_{i,t-i} \left(\mathbb{1}_{\text{SC}_t} + \frac{C_t - B_t}{R_t} \mathbb{1}_{\text{SC}_t^c \cap \text{SCB}_t} \right). \end{aligned} \quad (15)$$

Assumption 4.1. The stochastic system $\{(B_{i,j}, R_{i,j}, P_{i,j}, C_{i+j}) : i \geq i_0, j \geq 0\}$ satisfies (15).

From (3) and (15) it follows immediately that

$$\begin{aligned} \mathbb{E}[B_{i,t-i+1} \mid \mathcal{F}_t \vee \mathcal{G}_t] &= B_{i,t-i} \left(1 - \frac{C_t}{B_t}\right) \mathbb{1}_{\text{SCB}_t^c} \\ &\quad + R_{i,t-i} \left(\mathbb{1}_{\text{SCB}_t^c} + \left(1 - \frac{C_t - B_t}{R_t}\right) \mathbb{1}_{\text{SC}_t^c \cap \text{SCB}_t} \right). \end{aligned} \quad (16)$$

Note that recursion (5) (which holds regardless of the choice of procedure for processing claims) can be written

$$B_{t+1} = B_t \left(1 - \frac{C_t}{B_t}\right) \mathbb{1}_{\text{SCB}_t^c} + R_t \left(\mathbb{1}_{\text{SCB}_t^c} + \left(1 - \frac{C_t - B_t}{R_t}\right) \mathbb{1}_{\text{SC}_t^c \cap \text{SCB}_t} \right).$$

Note also that by summing over occurrence periods we obtain the same recursion from (16) since in stationarity

$$\sum_i \mathbb{E}[B_{i,t-i+1} \mid \mathcal{F}_t \vee \mathcal{G}_t] = \mathbb{E}[B_{t+1} \mid \mathcal{F}_t \vee \mathcal{G}_t] = B_{t+1}.$$

Hence, whereas (5) explains the backlog dynamics on the aggregate level, (16) adds information by explaining the backlog dynamics on an individual occurrence period level. We emphasize that (5) holds for any procedure for processing claims, whereas (16) is a consequence of a particular choice of the claims processing procedure. If, as an example, we would require that newly reported claims should be processed before backlog claims, then (5) would still hold, whereas (16) would result in another expression.

5 Computation of backlog expectations

The aim of this section is to provide explicit expressions for unconditional and conditional expectations of the number of backlog claims $B_{i,j}$. Considering the unconditional expectation $\mathbb{E}[B_{i,j}]$ makes most sense if we consider the backlog process in its stationary state. We will therefore (see Assumption 5.1) assume that (R_t) and (C_t) are i.i.d. sequences with $\mathbb{E}[R_t] < \mathbb{E}[C_t]$ which ensures a stationary distribution and that the Markov chain (B_t) approaches stationarity from an arbitrary fixed initial state. When the initial state B_0 of the backlog process $(B_t)_{t=0}^\infty$ is drawn from its stationary distribution, the distribution of $B_{i,j}$ does not depend on i . On the other hand, we will consider conditional expectations $\mathbb{E}[B_{\tau-i,i+k+1} \mid \mathcal{G}_\tau]$ for $k \geq 0$. For such conditional expectations, stationarity issues for (B_t) do not play a role, but the assumption of i.i.d. sequences (R_t) and (C_t) is imposed in order to obtain an explicit expression for the conditional expectation. Conditioning on \mathcal{G}_τ means that we consider a situation where $B_{i,j}$ and $R_{i,j}$ for $i+j \leq \tau$ are observable, e.g., this may address the question of optimally planing capacities when currently facing a large backlog B_τ . This corresponds to data likely available to an actuary.

Assumption 5.1. The stochastic system $\{(B_{i,j}, R_{i,j}, P_{i,j}, C_{i+j}) : i \geq i_0, j \geq 0\}$ satisfies:

- (i) (R_t) and (C_t) are i.i.d. sequences with $0 < \mathbb{E}[R_t] < \mathbb{E}[C_t] < \infty$.
- (ii) There exist constants $\mu_j > 0$ such that $\mathbb{E}[R_{i,j} | \mathcal{F}_{i+j}] = \mathbb{E}[R_{i,j} | R_{i+j}] = (\mu_j/\mu)R_{i+j}$ for all i , where $\mu = \sum_{j=0}^{J_R} \mu_j$.

Assumption 5.1 (ii) holds for a wide family of stochastic models. On a sufficiently rich probability space, the requirement is essentially that $R_{i,j}$ and $R_{i+j} - R_{i,j}$ can be represented as independent increments of a non-negative Lévy process. We refer to [14] for more on Lévy processes.

Theorem 5.2. *Assume that $R_{i,j}$ and $R_{i+j} - R_{i,j}$ are independent non-negative random variables with $\mathbb{E}[R_{i+j}] = \mu \in (0, \infty)$ and $\mathbb{E}[R_{i,j}] = \mu_j \in (0, \mu)$, and assume that there exists a Lévy process $(X_t)_{t \geq 0}$ such that $X_1 = R_{i+j}$ and $X_{\mu_j/\mu} = R_{i,j}$. Then $\mathbb{E}[R_{i,j} | R_{i+j}] = (\mu_j/\mu)R_{i+j}$.*

Proof of Theorem 5.2. First, note that for any i.i.d. random variables Z_1, \dots, Z_n with sum S_n , by symmetry it holds that $\mathbb{E}[Z_1 | S_n] = S_n/n$. Consider a sequence $((m_k, n_k))_{k \geq 1}$, with $(m_k, n_k) \in \mathbb{Z}^2$, such that $\lim_{k \rightarrow \infty} m_k/n_k = \mu_j/\mu$. For each k , let

$$Z_v^k := X_{v/n_k} - X_{(v-1)/n_k}, \quad v = 1, \dots, n_k, \quad S_u^k := \sum_{v=1}^u Z_v^k.$$

Note that

$$\mathbb{E}[S_{m_k}^k | S_{n_k}^k] = \frac{m_k}{n_k} S_{n_k}^k = \frac{m_k}{n_k} X_1,$$

and by the stochastic continuity property for Lévy processes

$$S_{m_k}^k \xrightarrow{\mathbb{P}} X_{\mu_j/\mu} \text{ as } k \rightarrow \infty.$$

Hence, there is a subsequence $k' \rightarrow \infty$ such that $S_{m_{k'}}^{k'} \xrightarrow{a.s.} X_{\mu_j/\mu}$. By Theorem 34.2(v) in [3],

$$\mathbb{E}[S_{m_{k'}}^{k'} | S_{n_{k'}}^{k'}] \xrightarrow{a.s.} \mathbb{E}[X_{\mu_j/\mu} | X_1],$$

from which the conclusion follows. \square

We now turn to the computation of backlog expectations. In order to avoid unnecessarily lengthy expressions we introduce the notation

$$F_t := R_t \left(\mathbb{1}_{\{B_t > C_t\}} + \left(1 - \frac{C_t - B_t}{R_t} \right) \mathbb{1}_{\{B_t \leq C_t < B_t + R_t\}} \right),$$

$$G_t := \left(1 - \frac{C_t}{B_t} \right) \mathbb{1}_{\{B_t > C_t\}},$$

and note that F_t and G_t are \mathcal{F}_t -measurable, see Section 4.

In Theorems 5.3 and 5.6 below we derive expressions for unconditional and conditional expectations of backlogs $B_{i,j}$. From these expressions we note that the expectations are fully determined once the corresponding expectations for quantities $F_{\tau+1}$, $F_{\tau}G_{\tau+1}$, $F_{\tau}G_{\tau+1}G_{\tau+2}$, etc., can be evaluated. We will consider numerical computation of the latter expectations in Section 7. We use the notation \wedge for the minimum $x \wedge y = \min(x, y)$.

Theorem 5.3. *Assume that Assumptions 2.1, 4.1 and 5.1 hold. For $j \geq 0$,*

$$\mathbb{E}[B_{i,j}] = \sum_{k=1}^{j \wedge (J+1)} \frac{\mu_{k-1}}{\mu} \mathbb{E} \left[F_{i+k-1} \prod_{l=k}^{j-1} G_{i+l} \right],$$

where an empty sum is equal to 0 and an empty product is equal to 1.

The identity $B_{i,j+1} = B_{i,j} + R_{i,j} - P_{i,j}$ implies the following expression for $\mathbb{E}[P_{i,j}]$:

Corollary 5.4. *Assume that Assumptions 2.1, 4.1 and 5.1 hold. For $j \geq 0$,*

$$\begin{aligned} \mathbb{E}[P_{i,j}] &= \sum_{k=1}^{j \wedge (J+1)} \frac{\mu_{k-1}}{\mu} \mathbb{E} \left[F_{i+k-1} \left(\prod_{l=k}^{j-1} G_{i+l} \right) (1 - G_{i+j}) \right] \\ &\quad + \mathbb{1}_{\{0 \leq j \leq J\}} \frac{\mu_j}{\mu} (\mu - \mathbb{E}[F_{i+j}]), \end{aligned}$$

where an empty sum is equal to 0 and an empty product is equal to 1.

In the proof of Theorem 5.3 we will use conditional independence properties together with the fact that products of the kind $G_{t+1} \cdots G_{t+h}$, $h \geq 1$, are $\sigma\{B_s, R_s, C_s : s > t\}$ -measurable.

Lemma 5.5. *Assume that Assumptions 2.1, 4.1 and 5.1 hold. Then \mathcal{H}_t and $\{B_s, R_s, C_s : s > t\}$ are conditionally independent given \mathcal{F}_t .*

Proof of Lemma 5.5. Note that

$$\sigma\{B_s, R_s, C_s : s > t\} = \sigma(B_{t+1}) \vee \sigma\{R_s, C_s : s > t\},$$

where $\sigma(B_{t+1}) \subset \mathcal{F}_t$ and $\sigma\{R_s, C_s : s > t\}$ are independent, and $\sigma\{R_s, C_s : s > t\}$ and \mathcal{H}_t are independent. Hence, \mathcal{H}_t and $\{B_s, R_s, C_s : s > t\}$ are dependent only through B_{t+1} , and therefore conditionally independent given \mathcal{F}_t . \square

Proof of Theorem 5.3. We prove the statement for $j = 0, 1, 2, 3$ in detail assuming $j \leq J + 1$. From this, it is obvious how to repeat the argument recursively to prove the statement for larger values of j . $B_{i,0} = 0$ by definition. For $j = 1$,

$$\begin{aligned} \mathbb{E}[B_{i,1}] &= \mathbb{E}[\mathbb{E}[B_{i,1} \mid \mathcal{F}_i \vee \mathcal{G}_i]] = \mathbb{E} \left[\frac{R_{i,0}}{R_i} F_i \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\frac{R_{i,0}}{R_i} F_i \mid \mathcal{F}_i \right] \right] = \mathbb{E} \left[\frac{\mathbb{E}[R_{i,0} \mid \mathcal{F}_i]}{R_i} F_i \right] = \frac{\mu_0}{\mu} \mathbb{E}[F_i]. \end{aligned}$$

For $j = 2$,

$$\mathbb{E}[B_{i,2}] = \mathbb{E}[\mathbb{E}[B_{i,2} \mid \mathcal{F}_{i+1} \vee \mathcal{G}_{i+1}]] = \mathbb{E}[B_{i,1}G_{i+1}] + \mathbb{E}\left[\frac{R_{i,1}}{R_{i+1}}F_{i+1}\right],$$

where for the newly reported claims

$$\mathbb{E}\left[\frac{R_{i,1}}{R_{i+1}}F_{i+1}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{R_{i,1}}{R_{i+1}}F_{i+1} \mid \mathcal{F}_{i+1}\right]\right] = \frac{\mu_1}{\mu}\mathbb{E}[F_{i+1}],$$

and for the previous backlog

$$\begin{aligned} \mathbb{E}[B_{i,1}G_{i+1}] &= \mathbb{E}[\mathbb{E}[B_{i,1}G_{i+1} \mid \mathcal{F}_i]] = \mathbb{E}[\mathbb{E}[B_{i,1} \mid \mathcal{F}_i]\mathbb{E}[G_{i+1} \mid \mathcal{F}_i]] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{E}[B_{i,1} \mid \mathcal{F}_i \vee \mathcal{G}_i] \mid \mathcal{F}_i]\mathbb{E}[G_{i+1} \mid \mathcal{F}_i]] \\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{R_{i,0}}{R_i}F_i \mid \mathcal{F}_i\right]\mathbb{E}\left[G_{i+1} \mid \mathcal{F}_i\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{R_{i,0}}{R_i} \mid \mathcal{F}_i\right]\mathbb{E}\left[F_iG_{i+1} \mid \mathcal{F}_i\right]\right] = \frac{\mu_0}{\mu}\mathbb{E}[F_iG_{i+1}], \end{aligned} \tag{17}$$

where we used Lemma 5.5 noting that $B_{i,1}$ is \mathcal{H}_i -measurable. For $j = 3$,

$$\mathbb{E}[B_{i,3}] = \mathbb{E}[\mathbb{E}[B_{i,3} \mid \mathcal{F}_{i+2} \vee \mathcal{G}_{i+2}]] = \mathbb{E}[B_{i,2}G_{i+2}] + \mathbb{E}\left[\frac{R_{i,2}}{R_{i+2}}F_{i+2}\right],$$

where

$$\mathbb{E}\left[\frac{R_{i,2}}{R_{i+2}}F_{i+2}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{R_{i,2}}{R_{i+2}}F_{i+2} \mid \mathcal{F}_{i+2}\right]\right] = \frac{\mu_2}{\mu}\mathbb{E}[F_{i+2}],$$

and, similarly to (17),

$$\begin{aligned} \mathbb{E}[B_{i,2}G_{i+2}] &= \mathbb{E}[\mathbb{E}[\mathbb{E}[B_{i,2} \mid \mathcal{F}_{i+1} \vee \mathcal{G}_{i+1}] \mid \mathcal{F}_{i+1}]\mathbb{E}[G_{i+2} \mid \mathcal{F}_{i+1}]] \\ &= \mathbb{E}\left[\mathbb{E}\left[B_{i,1}G_{i+1} + \frac{R_{i,1}}{R_{i+1}}F_{i+1} \mid \mathcal{F}_{i+1}\right]\mathbb{E}[G_{i+2} \mid \mathcal{F}_{i+1}]\right] \\ &= \mathbb{E}[B_{i,1}G_{i+1}G_{i+2}] + \mathbb{E}\left[\frac{R_{i,1}}{R_{i+1}}F_{i+1}G_{i+2}\right], \end{aligned}$$

where we used Lemma 5.5. The last two terms are simplified as follows

$$\begin{aligned} \mathbb{E}\left[\frac{R_{i,1}}{R_{i+1}}F_{i+1}G_{i+2}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{R_{i,1}}{R_{i+1}}F_{i+1}G_{i+2} \mid \mathcal{F}_{i+1}\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{R_{i,1}}{R_{i+1}}F_{i+1} \mid \mathcal{F}_{i+1}\right]\mathbb{E}\left[G_{i+2} \mid \mathcal{F}_{i+1}\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{R_{i,1}}{R_{i+1}} \mid \mathcal{F}_{i+1}\right]\mathbb{E}\left[F_{i+1}G_{i+2} \mid \mathcal{F}_{i+1}\right]\right] \\ &= \frac{\mu_1}{\mu}\mathbb{E}[F_{i+1}G_{i+2}], \end{aligned}$$

where we used Lemma 5.5. For the remaining term we recall that $B_{i,1}$ is \mathcal{H}_i -measurable and use Lemma 5.5 to receive

$$\begin{aligned}
\mathbb{E}[B_{i,1}G_{i+1}G_{i+2}] &= \mathbb{E}[\mathbb{E}[B_{i,1}G_{i+1}G_{i+2} \mid \mathcal{F}_i]] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[B_{i,1} \mid \mathcal{F}_i \vee \mathcal{G}_i] \mid \mathcal{F}_i] \mathbb{E}[G_{i+1}G_{i+2} \mid \mathcal{F}_i]] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{R_{i,0}}{R_i}F_i \mid \mathcal{F}_i\right] \mathbb{E}[G_{i+1}G_{i+2} \mid \mathcal{F}_i]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{R_{i,0}}{R_i} \mid \mathcal{F}_i\right] \mathbb{E}[F_iG_{i+1}G_{i+2} \mid \mathcal{F}_i]\right] \\
&= \frac{\mu_0}{\mu} \mathbb{E}[F_iG_{i+1}G_{i+2}].
\end{aligned}$$

For $j > 3$ the statement follows by reusing the above arguments. \square

Theorem 5.3 considered unconditional backlog expectations. Below follows the corresponding result for conditional expectations.

Theorem 5.6. *Assume that Assumptions 2.1, 4.1 and 5.1 hold. For $k \geq 0$ such that $\tau - i + k \geq 0$,*

$$\begin{aligned}
&\mathbb{E}[B_{i,\tau-i+k+1} \mid \mathcal{G}_\tau] \\
&= \sum_{m=1}^k \mathbb{1}_{\{i \leq \tau+m \leq i+J\}} \frac{\mu_{\tau-i+m}}{\mu} \mathbb{E}\left[F_{\tau+m} \prod_{l=m+1}^k G_{\tau+l} \mid \mathcal{E}_\tau\right] \quad (18)
\end{aligned}$$

$$+ \mathbb{1}_{\{i \leq \tau \leq i+J\}} \frac{R_{i,\tau-i}}{R_\tau} \mathbb{E}[F_\tau \mid \mathcal{E}_\tau] \mathbb{E}\left[\prod_{l=1}^k G_{\tau+l} \mid \mathcal{E}_\tau\right] \quad (19)$$

$$+ \mathbb{1}_{\{i \leq \tau\}} B_{i,\tau-i} \mathbb{E}[G_\tau \mid \mathcal{E}_\tau] \mathbb{E}\left[\prod_{l=1}^k G_{\tau+l} \mid \mathcal{E}_\tau\right], \quad (20)$$

where an empty sum is equal to 0 and an empty product is equal to 1. We can equally replace all conditions \mathcal{E}_τ by \mathcal{G}_τ .

Proof of Theorem 5.6. The statement is proved by the same arguments as in the proof of Theorem 5.3. \square

Remark 5.7. For past occurrence periods, $i \leq \tau$, all three terms (18), (19), (20) contribute to the conditional backlog expectation. For future occurrence periods, $i > \tau$, the terms (19) and (20) vanish.

Remark 5.8. If $C_t = c$ is constant, then F_t and G_t are fully determined by B_t and R_t . Hence, $\mathbb{E}[F_\tau \mid \mathcal{E}_\tau] = F_\tau$ and $\mathbb{E}[G_\tau \mid \mathcal{E}_\tau] = G_\tau$. Moreover, since (R_t) is an i.i.d. sequence, if $C_t = c$ is constant, then B_{t+1} is fully determined by B_t

and R_t , and

$$\begin{aligned}\mathbb{E}\left[F_{\tau+m} \prod_{l=m+1}^k G_{\tau+l} \mid \mathcal{E}_\tau\right] &= \mathbb{E}\left[F_{\tau+m} \prod_{l=m+1}^k G_{\tau+l} \mid B_{\tau+1}\right], \\ \mathbb{E}\left[\prod_{l=1}^k G_{\tau+l} \mid \mathcal{E}_\tau\right] &= \mathbb{E}\left[\prod_{l=1}^k G_{\tau+l} \mid B_{\tau+1}\right].\end{aligned}$$

Remark 5.8 implies the following corollary to Theorem 5.6.

Corollary 5.9. *Assume that Assumptions 2.1, 4.1 and 5.1 hold. Assume that there exists a non-random $c > 0$ such that $C_t = c$ for all t . For $k \geq 0$ such that $\tau - i + k \geq 0$,*

$$\begin{aligned}\mathbb{E}[B_{i,\tau-i+k+1} \mid \mathcal{G}_\tau] &= \sum_{m=1}^k \mathbb{1}_{\{i \leq \tau+m \leq i+J\}} \frac{\mu_{\tau-i+m}}{\mu} \mathbb{E}\left[F_{\tau+m} \prod_{l=m+1}^k G_{\tau+l} \mid B_{\tau+1}\right] \\ &\quad + \mathbb{1}_{\{i \leq \tau \leq i+J\}} \frac{R_{i,\tau-i}}{R_\tau} F_\tau \mathbb{E}\left[\prod_{l=1}^k G_{\tau+l} \mid B_{\tau+1}\right] \\ &\quad + \mathbb{1}_{\{i \leq \tau\}} B_{i,\tau-i} G_\tau \mathbb{E}\left[\prod_{l=1}^k G_{\tau+l} \mid B_{\tau+1}\right],\end{aligned}$$

where an empty sum is equal to 0 and an empty product is equal to 1.

6 The negative binomial model

It may seem natural to consider a Poisson model for the number of reported claims, where all $R_{i,j}$ are independent and $R_{i,j} \sim \text{Pois}(\mu_j)$. Such a model is consistent with Assumption 5.1. However, as explained in Remark 2.3, for constant capacity $C_t = c$ and $\mu < c$ reasonably large, say 1000, the expected total size of the backlog will be close to 0 unless $c \approx \mu$. The event $\{B_t > c\}$, which appears in expected backlog calculations in Section 5, is a very unlikely event for the Poisson model. Markov's inequality together with Remark 2.3 give

$$\mathbb{P}[B_t > c] \leq \frac{1}{c/\mu - 1} \frac{1}{2c} \approx 0.$$

Thus, long-term backlogs are not an issue in Poisson settings. We need a model for the numbers of reported claims with considerable over-dispersion compared to the Poisson model to have an interesting backlog behavior.

We start with a generic negative binomial random variable R . Assume that R is conditionally Poisson distributed with conditional mean Λ , and $\Lambda \sim \Gamma(\alpha, \beta)$. It follows that R has an unconditional negative binomial distribution $\text{NegBin}(\alpha, \beta)$ with moment generating function

$$\mathbb{E}[\exp\{xR\}] = \mathbb{E}[\exp\{\Lambda(e^x - 1)\}] = \left(\frac{\beta}{\beta - (e^x - 1)}\right)^\alpha, \quad \text{for } x < \log(\beta + 1).$$

From this it follows that we can aggregate independent negative binomial random variables $R_{i,0}, \dots, R_{i,J}$ (or $R_{t,0}, R_{t-1,1}, \dots, R_{t-J,J}$) as long as they share the same scale parameter β , and we remain in the family of negative binomial random variables

$$\mathbb{E} \left[\exp \left\{ x \sum_{j=0}^J R_{i,j} \right\} \right] = \prod_{j=0}^J \mathbb{E}[\exp\{xR_{i,j}\}] = \left(\frac{\beta}{\beta - (e^x - 1)} \right)^{\sum_{j=0}^J \alpha_j}.$$

We consider the following model for the number of reported claims: all $R_{i,j}$ are independent with $R_{i,j} \sim \text{NegBin}(\alpha_j, \beta)$. Hence,

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

with means and variances

$$\mathbb{E}[R_t] = \alpha/\beta \quad \text{and} \quad \text{Var}[R_t] = \alpha/\beta + \alpha/\beta^2 = \mathbb{E}[R_t](1 + 1/\beta).$$

The over-dispersion term $1/\beta$ distinguishes the negative binomial model from the Poisson model. In practical applications, it is often reasonable to assume that the model has a coefficient of variation roughly on the unit scale, typically, it is bigger for claim size modeling than for claim counts modeling. The coefficient of variation is in this negative binomial model given by

$$\frac{\sqrt{\text{Var}[R_t]}}{\mathbb{E}[R_t]} = \sqrt{\frac{\beta}{\alpha} + \frac{1}{\alpha}} = \sqrt{\frac{1}{\mathbb{E}[R_t]} + \frac{1/\beta}{\mathbb{E}[R_t]}}.$$

This requires that $\beta > 0$ lives on the same scale as $1/\mathbb{E}[R_t]$. We select $\alpha = 2$ and $\beta = 2/1000$, this gives an expected value of $\mu = \mathbb{E}[R_t] = 1000$ and a coefficient of variation ≈ 0.7 . We select $J = 3$, and we split the expected numbers of reported claims according to

$$\begin{aligned} (\mathbb{E}[R_{t,0}], \mathbb{E}[R_{t,1}], \mathbb{E}[R_{t,2}], \mathbb{E}[R_{t,3}]) &= (\alpha_0, \alpha_1, \alpha_2, \alpha_3)/\beta \\ &= (500, 300, 150, 50). \end{aligned} \tag{21}$$

We select a constant period capacity $C_t = c$ with $c = \eta\mu$, where the capacity ratio η is chosen as $\eta = 1.2$.

The following figures show the results from this negative binomially generated claims reportings data R_t , $1 \leq t \leq T$, over the $T = 120$ (monthly) time periods. We start the process with a zero backlog $B_1 = 0$.

Figure 1 (lhs) shows the probabilities of a non-zero backlog when starting with a zero backlog, $\mathbb{P}[B_t > 0 \mid B_1 = 0]$, for different time periods $1 \leq t \leq T$, and the corresponding conditionally expected backlog sizes $\mathbb{E}[B_t \mid B_t > 0, B_1 = 0]$ for a capacity ratio of $\eta = 1.2$ in this negative binomial model. The conditionally expected long-term backlog $\lim_{t \rightarrow \infty} \mathbb{E}[B_t \mid B_t > 0, B_1 = 0]$ is of magnitude 1600, see Figure 1 (rhs). This implies that we have an expected long-term backlog of

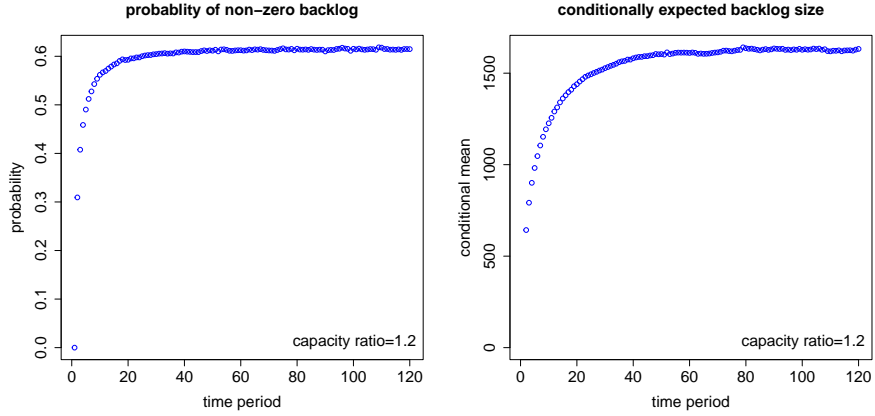


Figure 1: Negative binomial model with capacity ratio $\eta = 1.2$: (lhs) probabilities of a non-zero backlog $\mathbb{P}[B_t > 0 \mid B_1 = 0]$ for different time periods $1 \leq t \leq T$; (rhs) conditionally expected backlog sizes $\mathbb{E}[B_t \mid B_t > 0, B_1 = 0]$.

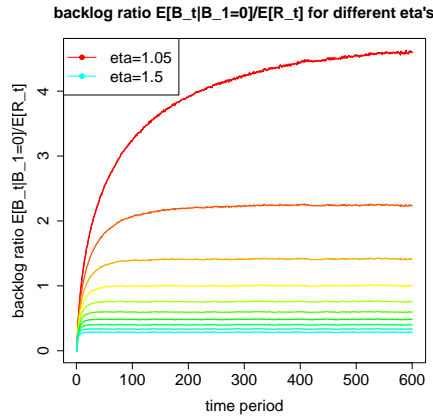


Figure 2: Relative expected backlogs $\mathbb{E}[B_t \mid B_1 = 0]/\mathbb{E}[R_t]$ for capacity ratios $\eta \in \{1.05, 1.10, \dots, 1.50\}$ if one starts with a zero backlog.

$\lim_{t \rightarrow \infty} \mathbb{E}[B_t \mid B_1 = 0] \approx 1000$, which is just slightly below the selected capacity $c = \eta\mu = 1200$. This results in frequent carry forward of old backlogs.

Figure 2 shows the relative expected backlogs $\mathbb{E}[B_t \mid B_1 = 0]/\mathbb{E}[R_t]$ if we start the process with a zero backlog $B_1 = 0$. The different colors correspond to capacity ratios $\eta \in \{1.05, 1.10, \dots, 1.50\}$. For a capacity ratio $\eta = 1.1$ the average backlog is roughly twice the expected number of reported claims, and for $\eta = 1.2$ we have a factor of roughly 1.

Figure 3 (lhs) gives the empirical conditional densities of the backlog B_t , conditioned on $\{B_t > 0, B_1 = 0\}$, for $2 \leq t \leq T$. The light-blue density with the highest maximum corresponds to $t = 2$, and with increasing time t we follow

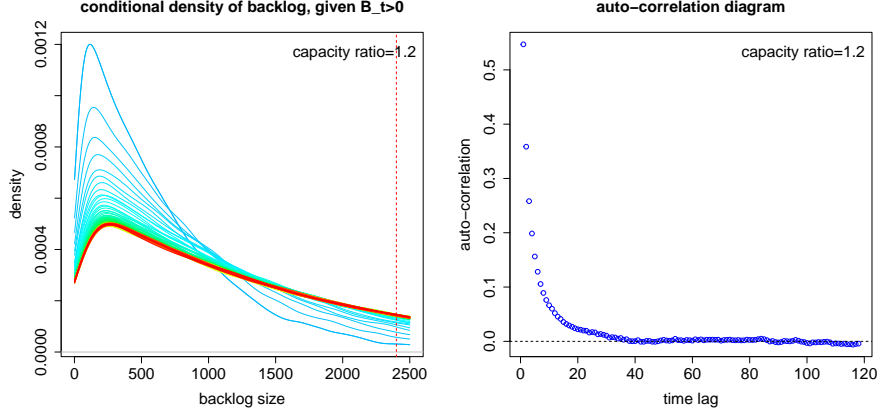


Figure 3: Negative binomial model: (lhs) empirical densities of the backlogs, conditioned on $\{B_t > 0, B_1 = 0\}$ for different time periods $2 \leq t \leq T$; (rhs) auto-correlation diagram of backlogs B_2 and B_{2+s} for time lags $s \geq 1$ when starting with a zero backlog $B_1 = 0$.

the rainbow colors (from light-blue to red). We observe that this conditional backlog size has still a significant positive probability that exceeds twice the capacity $2c$ in the stationary limit (vertical red dotted line). This indicates that we likely have carry forwards of backlogs over multiple periods. Figure 3 (rhs) shows the auto-correlation diagram of backlogs B_2 and B_{2+s} for time lags $s \geq 1$ when starting with a zero backlog $B_1 = 0$. This auto-correlation $\text{Corr}(B_2, B_{2+s} | B_1 = 0)$ has vanished after 40 periods.

Figure 4 (lhs) shows the empirical density of F_t and the conditional empirical density of F_t , given $B_t = 0$. The latter is directly simulated, given $B_t = 0$, using the aggregate negative binomial model assumption for R_t . For the former, we select B_t from the stationary limit distribution of Figure 3 (lhs), and then simulate F_t conditional on this value. The vertical dotted lines show the (conditionally) expected values $\mathbb{E}[F_t]$ and $\mathbb{E}[F_t | B_t = 0]$ in red and blue.

This carries over to Figure 4 (rhs) which shows the (conditional) expectations g_j and $g_j(0)$ of $F_t \prod_{l=t+1}^{t+j} G_l$, where

$$g_j := \mathbb{E} \left[F_t \prod_{l=t+1}^{t+j} G_l \right] \quad \text{and} \quad g_j(b) := \mathbb{E} \left[F_t \prod_{l=t+1}^{t+j} G_l \mid B_t = b \right]. \quad (22)$$

We obtain expected values that are significantly bigger than zero for small j 's, and this precisely differentiates the backlog behavior in the negative binomial model from the Poisson case.

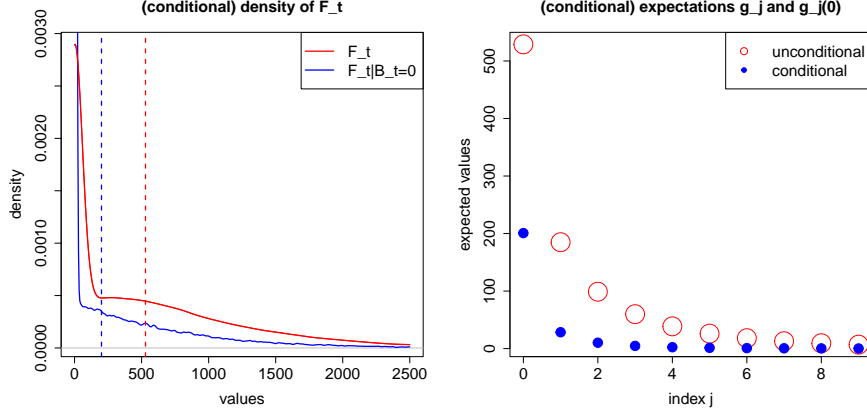


Figure 4: Negative binomial model: (lhs) (conditionally) densities and expected values $\mathbb{E}[F_t]$ and $\mathbb{E}[F_t | B_t = 0]$; (rhs) (conditional) expectations given in equation (22) for $0 \leq j \leq 9$, where $j = 0$ corresponds to the empty product only considering F_t ; red refers to the unconditional case averaging over the stationary distribution for the initial value B_t , and blue corresponds to starting the process in a zero backlog $B_t = 0$.

7 Neural network approximation

To take into account backlog costs in cost optimization problems, we need to be able to study the sensitivities of the functions $(g_j(\cdot))_{j \geq 0}$ defined in (22) in the capacity ratio parameter $\eta > 1$. We therefore slightly modify the corresponding notation by adding upper indices (η) to the variables that directly depend on the selected capacity ratio $\eta > 1$. In particular, we consider the following quantity for backlog computations

$$g_j(b; \eta) = \mathbb{E} \left[F_1^{(\eta)} \prod_{l=2}^{j+1} G_l^{(\eta)} \mid B_1^{(\eta)} = b \right], \quad (23)$$

for $j \geq 0$, with an empty product being set to one.

Since the quantities (23) cannot be computed explicitly, as a function of η , we fit a recurrent neural network (RNN) $\mathbf{z}_\theta^{\text{RNN}} : \mathbb{R}^2 \rightarrow \mathbb{R}^T$ that takes the inputs (b, η) to approximate the function $(b, \eta) \mapsto (g_j(b; \eta))_{j=0}^{T-1} \in \mathbb{R}^T$, for a fixed large T . The fitted RNN $\mathbf{z}_\theta^{\text{RNN}}$ is obtained by minimizing the square loss in network parameter θ

$$\hat{\theta} \in \arg \min_{\theta} \frac{1}{n} \sum_{k=1}^n \sum_{j=0}^{T-1} \left(F_1^k \prod_{l=2}^{j+1} G_l^k - \mathbf{z}_\theta^{\text{RNN}}(B_1^k, \eta^k)_j \right)^2, \quad (24)$$

where the observations $(F_1^k, G_2^k, \dots, G_{T-1}^k)$, $1 \leq k \leq n$, are simulated by using i.i.d. randomized uniform capacity ratios $\eta^k \in (1.05, 1.50)$, randomized initial

backlogs B_1^k (from the stationary limit distribution corresponding to the simulated η^k), simulated reporting processes $(R_t^k)_{t=1}^T$, and the resulting backlog processes $(B_t^k)_{t=1}^T$ for the simulated capacity ratios η^k . From this we compute the observations $(F_1^k, G_2^k, \dots, G_{T-1}^k)$, which then enter the square loss (24).

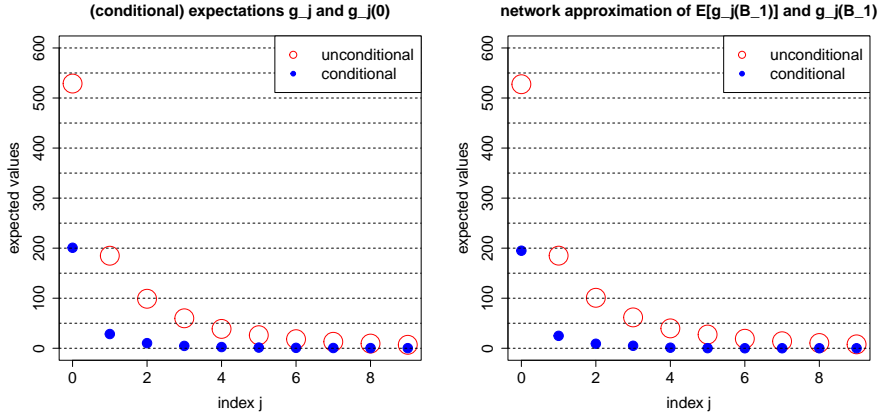


Figure 5: (Conditional) expectations $\mathbb{E}[g_j(B_1^{(\eta)}; \eta)]$ and $g_j(0; \eta)$, $0 \leq j \leq 9$, for capacity ratio $\eta = 1.2$: (lhs) empirical means taken from Figure 4 (rhs), and (rhs) RNN approximations $z_{\hat{\theta}}^{\text{RNN}}$.

Figure 5 (rhs) shows the results of the fitted RNN $z_{\hat{\theta}}^{\text{RNN}}$ approximation, and they are compared to the empirical means on the left-hand side in this figure; these are taken from Figure 4 (rhs), and both figures have the same y -scale and the same selected capacity ratio $\eta = 1.2$. For the conditional version we start with a zero backlog $B_1^{(\eta)} = 0$, thus, we consider $z_{\hat{\theta}}^{\text{RNN}}(0, \eta)$, and for the unconditional version $\mathbb{E}[z_{\hat{\theta}}^{\text{RNN}}(B_1^{(\eta)}, \eta)]$ we average over the stationary limit distribution for $B_1^{(\eta)}$ that corresponds to the capacity ratio $\eta = 1.2$; note that this averaging is done purely empirically by selecting stationary samples from the backlogs. From this figure we conclude that the RNN approximations are very close to the empirical means. Thus, overall the RNN approximations $z_{\hat{\theta}}^{\text{RNN}}$ seem very accurate.

Figure 6 shows the RNN approximations for different choices of the capacity ratio $\eta \in \{1.05, 1.10, \dots, 1.50\}$ and for different starting backlogs $B_1^{(\eta)}$. All these plots are obtained from the (single) fitted RNN $(b, \eta) \mapsto z_{\hat{\theta}}^{\text{RNN}}(b, \eta)$, i.e., we can now simultaneously evaluate $g_j(b, \eta)$ for any initial backlog $b \in \{0, \dots, 40,000\}$ and any capacity ratio $\eta \in [1.05, 1.50]$, this is the input domain on which the network $z_{\hat{\theta}}^{\text{RNN}}$ has been trained on.

Figure 6 (top-left) gives the unconditional versions $\mathbb{E}[z_{\hat{\theta}}^{\text{RNN}}(B_1^{(\eta)}; \eta)]$, where we average over the stationary limit distribution of the backlogs under the given capacity ratio η ; note that this averaging is again done purely empirically by selecting stationary samples from the backlogs. Figure 6 (top-right) shows the

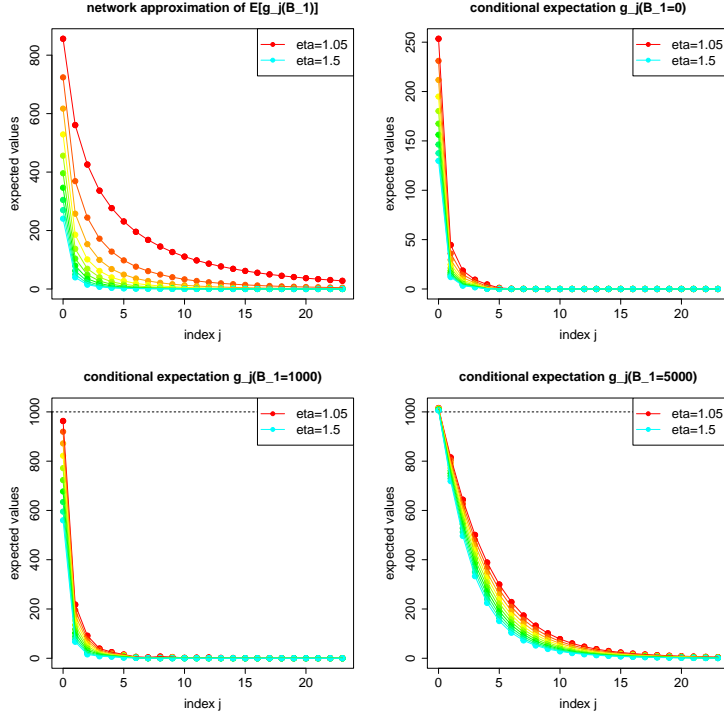


Figure 6: RNN approximations of the means $\mathbb{E}[g_j(B_1^{(\eta)}; \eta)]$ and conditional means $g_j(B_1^{(\eta)}; \eta)$ for capacity ratios $\eta \in \{1.05, 1.10, \dots, 1.50\}$: (top-left) unconditional version averaged over the stationary limit distribution of $B_1^{(\eta)}$; remaining versions are the conditional means for starting backlogs $B_1^{(\eta)} \in \{0, 1000, 5000\}$.

conditional versions $z_{\hat{\theta}}^{\text{RNN}}(0; \eta)$ starting in a zero backlog. The remaining plots show the conditional versions for starting backlogs $B_1^{(\eta)} \in \{1000, 5000\}$; these additional plots have all identical y -scales and the horizontal dotted lines is at the expected number of reported claims level $\mu = \mathbb{E}[R_1] = 1000$. Recall the random variable

$$F_1^{(\eta)} = R_1 \left(\mathbb{1}_{\{B_1^{(\eta)} > c^{(\eta)}\}} + \left(1 - \frac{c^{(\eta)} - B_1^{(\eta)}}{R_1} \right) \mathbb{1}_{\{B_1^{(\eta)} \leq c^{(\eta)} < B_1^{(\eta)} + R_1\}} \right).$$

Clearly, the first indicator is zero for $B_1^{(\eta)} = 1000$ and $c^{(\eta)} = \eta\mu = \eta 1000 > 1000$. Thus, only the second indicator contributes to $g_0(b, \eta)$ for $b < \eta\mu$. On the other hand, for any starting backlog $B_1^{(\eta)} \geq \eta\mu$ we have a non-vanishing first indicator, saying that $g_0(b, \eta) \geq \mu = \mathbb{E}[R_1]$ for $b \geq \eta\mu$. This is how the initial values $g_0(b, \eta)$ of Figure 6 (bottom) are interpreted, and this is highlighted by the horizontal darkgray dotted line. For $j \geq 1$, we can then (simply) verify that it takes more

time to run off the initial backlog $B_1^{(\eta)}$ for smaller capacity ratios $\eta > 1$.

Fitting the RNN z_θ^{RNN} becomes increasingly difficult the closer the capacity ratio $\eta > 1$ approaches its limit 1 of a non-explosive model which requires $c^{(\eta)} > \mu = \mathbb{E}[R_1]$. The unconditional version Figure 6 (top-left) is more sensitive in η than its conditional counterparts $g_0(b, \eta)$, because for the former we average the initial backlog $B_1^{(\eta)}$ over its stationary limit distribution, and the additional variability enters through the increasing volatility of this limiting distribution for decreasing η , i.e., we have slower convergence to the stationary limit distribution of the backlog process the closer $\eta > 1$ is to the critical value of one.

7.1 Approximation of unconditional backlog expectations

In view of the unconditional cost allocation problem, see Section 3, the conditional version $g_j(b; \eta)$, given in (23), is not fully suitable because it still needs averaging over the stationary limit distribution of the backlog, which can only be done numerically; this is exactly how Figure 6 (top-left) has been obtained. Here we would rather like to have a function

$$\eta \mapsto g_j(\eta) = \mathbb{E}\left[g_j(B_1^{(\eta)}; \eta)\right] = \mathbb{E}\left[F_1^{(\eta)} \prod_{l=2}^{j+1} G_l^{(\eta)}\right] \quad \text{for } j \geq 0. \quad (25)$$

We fit a different second RNN to directly approximate $\eta \mapsto (g_j(\eta))_{j=0}^{T-1}$ rather than $(b, \eta) \mapsto (g_j(b, \eta))_{j=0}^{T-1}$. For receiving a suitable RNN approximation we need to ensure that the initial backlog $B_1^{(\eta)}$ is sampled from the stationary limit distribution. For this we start a process that has some burn-in until it generates stationary samples. We first study empirically the convergence rate to the stationary phase. Figure 7 shows the results for two different capacity ratios $\eta = 1.05, 1.10$ starting from a zero backlog. From these plots we conclude that we need to simulate roughly 1200 iteration steps to arrive at an empirical approximation to the stationary limit distribution. All following results are obtained by using this burn-in of 1200 iterations, and the subsequent samples are then taken as an empirical approximation to the stationary limit distribution.

Based on this sampling and fitting strategy we fit a single input RNN to the function in (25). Figure 8 (rhs) shows the results of this direct fitting of a RNN to the unconditional means $g_j(\eta)$ using stationary backlog time series for different capacity ratios $\eta \in [1.05, 1.50]$. This is compared to the empirical average over the conditional means $g_j(B_1^{(\eta)}; \eta)$ taken from Figure 6 (top-left). We see an excellent alignment of the results, telling that both networks have learned the same structure. The only (smaller) differences are visible for the smallest capacity ratio $\eta = 1.05$. The issue here clearly is slow convergence to the stationary limiting distribution, which is also verified from Figure 7. For the unconditional cost optimization problem we use the unconditional network approximation to $(g_j(\eta))_{j \geq 0}$.

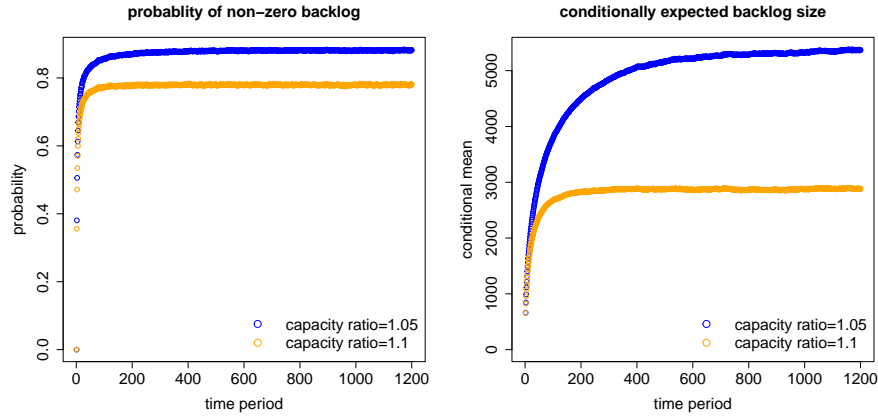


Figure 7: Negative binomial model with capacity ratios $\eta = 1.05, 1.10$: analysis of burn-in to reach the stationary limit distribution for an initial backlog of zero.

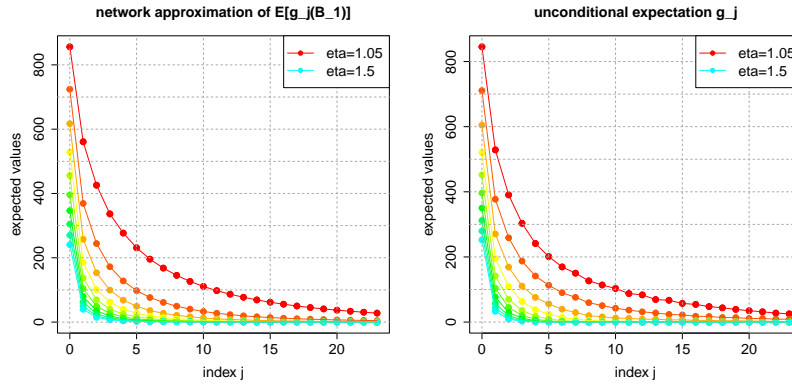


Figure 8: RNN approximations of the means $(g_j(\eta))_{j \geq 0}$ for capacity ratios $\eta \in \{1.05, 1.10, \dots, 1.50\}$: (lhs) this is identical to Figure 6 (top-left) received by empirically averaging over the conditional networks, and (rhs) direct fit of the unconditional mean using a single input network.

7.2 Approximation of conditional backlog expectations

In order to numerically approximate the conditional backlog expectations in Corollary 5.9 we observe that we need to have time-delayed versions of the quantities $g_j(b; \eta)$ introduced in (23). We need to compute $h_j(b, m; \eta)$, where,

for $j \geq 0$,

$$h_j(b, 0; \eta) := \mathbb{E} \left[\prod_{l=1}^j G_l^{(\eta)} \mid B_1^{(\eta)} = b \right], \quad (26)$$

$$h_j(b, m; \eta) := \mathbb{E} \left[F_m^{(\eta)} \prod_{l=m+1}^{m+j} G_l^{(\eta)} \mid B_1^{(\eta)} = b \right], \quad m \geq 1. \quad (27)$$

The functions $h_j(b, 0; \eta)$ are time-advanced versions of $g_j(b; \eta)$ by dropping $F_1^{(\eta)}$, the functions $h_j(b, m; \eta)$, $m \geq 1$, are time-delayed versions of $g_j(b; \eta)$. Using these we rewrite the expression for the conditional expectation in Corollary 5.9 as follows. For $k \geq 0$ such that $\tau - i + k \geq 0$,

$$\mathbb{E} [B_{i, \tau-i+k+1}^{(\eta)} \mid \mathcal{G}_\tau^{(\eta)}] = \sum_{m=1}^k \mathbb{1}_{\{i \leq \tau+m \leq i+J\}} \frac{\mu_{\tau-i+m}}{\mu} h_{k-m}(B_{\tau+1}^{(\eta)}, m; \eta) \quad (28)$$

$$+ \mathbb{1}_{\{i \leq \tau \leq i+J\}} \frac{R_{i, \tau-i}}{R_\tau} F_\tau^{(\eta)} h_k(B_{\tau+1}^{(\eta)}, 0; \eta) \quad (29)$$

$$+ \mathbb{1}_{\{i \leq \tau\}} B_{i, \tau-i}^{(\eta)} G_\tau^{(\eta)} h_k(B_{\tau+1}^{(\eta)}, 0; \eta). \quad (30)$$

Since $h_j(b, m; \eta)$ lives on different scales for $m = 0$, formula (26), and $m \geq 1$, formula (27), we fit two different networks to these two cases. This is easier in training. The case $m = 0$ is then completely analogous to $g_j(b; \eta)$, see (24), and the fitting results are given in Figure 12 in the appendix.

Fitting $h_j(b, m; \eta)$, $m \geq 1$, is slightly more tricky because the first term $F_1^{(\eta)}$ lives on a different scale compared to $G_l^{(\eta)}$, $l \geq m+1$. This is similar to $g_j(b; \eta)$. Note that we can rewrite (27) as follows

$$\begin{aligned} h_j(b, m; \eta) &= \mathbb{E} \left[F_m^{(\eta)} \prod_{l=m+1}^{m+j} G_l^{(\eta)} \mid B_1^{(\eta)} = b \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[F_m^{(\eta)} \prod_{l=m+1}^{m+j} G_l^{(\eta)} \mid B_m^{(\eta)}, B_1^{(\eta)} = b \right] \mid B_1^{(\eta)} = b \right] \\ &= \mathbb{E} \left[g_j(B_m^{(\eta)}; \eta) \mid B_1^{(\eta)} = b \right]. \end{aligned}$$

This shows that we can employ a similar approximation and fitting strategy as in (24) but we need to time-delay by $m-1$ periods. This requires an additional input m to the RNN $\mathbf{z}_\theta^{\text{RNN}}$, and then we can fit

$$\hat{\theta} \in \arg \min_{\theta} \frac{1}{n} \sum_{k=1}^n \sum_{j=0}^{T-1} \left(F_1^{k,m} \prod_{l=2}^{j+1} G_l^{k,m} - \mathbf{z}_\theta^{\text{RNN}}(B_1^k, \eta^k, m)_j \right)^2, \quad (31)$$

where the observations $(F_1^{k,m}, G_2^{k,m}, \dots, G_{T-1}^{k,m})$, $1 \leq k \leq n$, are received by first simulating the $m-1$ periods delayed starting point B_m^k from B_1^k , set new

starting value $B_1^{k,m} := B_m^k$, and then proceed as in (24). This provides us with a fitted neural network $(b, \eta, m) \mapsto z_{\hat{\theta}}^{\text{RNN}}(b, \eta, m)$ that approximates $h_j(b, m; \eta)$, $m \geq 1$, $j \geq 0$, $b \geq 0$ and $\eta \in (1.05, 1.50)$. The results are shown in Figures 13-15 in the appendix for the three different starting values $b \in \{0, 1000, 5000\}$.

8 Cost optimization

We now have prepared all the necessary numerical tools to study the optimal capacity ratio $\eta > 1$ for obtaining minimal costs. We distinguish the unconditional and the conditional cases. The latter considers a situation where we want to optimally plan the capacity under a given starting backlog $B_1^{(\eta)}$, and the former unconditional case is a global consideration of long-term optimal planning to receive minimal costs. In the long run, the conditional version will converge to the unconditional one, regardless of the specific initial backlog $B_1^{(\eta)}$.

8.1 Unconditional cost optimization

Recall the expressions (8) and (9) for expected combined delay-adjusted claim costs and capacity costs. Here we study these expected costs as functions of the capacity ratio η . We write

$$\mu_i^{(\ell)}(\eta) := \kappa_g \mu + \kappa_b \mathbb{E}[B^{(\eta)}] + \kappa_c (c^{(\eta)} - \mu) \quad (32)$$

for the linear cost model, and

$$\mu_i^{(\iota)}(\eta) := \kappa_g \sum_{j \geq 0} \lambda_b^j \mathbb{E}[B_{i,j}^{(\eta)} + R_{i,j} - B_{i,j+1}^{(\eta)}] + \kappa_c (c^{(\eta)} - \mu) \quad (33)$$

for the model with non-linear delay-inflated costs. We study these total costs as a function of the capacity ratio $\eta > 1$ providing $c^{(\eta)} = \eta\mu$ and impacting the backlog via recursion (5). Using the network approximation to $(g_j(\eta))_{j \geq 0}$ we can explicitly compute these costs using relation (see Theorem 5.3)

$$\mathbb{E}[B_{i,j}^{(\eta)}] = \sum_{k=1}^{j \wedge (J+1)} \frac{\mu_{k-1}}{\mu} \mathbb{E} \left[F_{i+k-1}^{(\eta)} \prod_{l=k}^{j-1} G_{i+l}^{(\eta)} \right] = \sum_{k=1}^{j \wedge (J+1)} \frac{\mu_{k-1}}{\mu} g_{j-k}(\eta), \quad (34)$$

where an empty sum is equal to 0 and an empty product is equal to 1.

Figure 9 (lhs) shows the optimal capacity ratio in the linear capacity cost case (32) with $\kappa_g = 1$, which gives ground-up costs of $\kappa_g \mu = 1000$. To this we add backlog costs with $\kappa_b = 0.075$ and excess capacity costs with $\kappa_c = 0.5$. The plot shows the linear cost case $\eta \mapsto \mu_i^{(\ell)}(\eta)$ as a function of the capacity ratio $\eta > 1$. In this model (and parametrization) the optimal (total cost minimizing) ratio is $\eta^* = 1.203$. Thus, in this case we have ground-up costs of 1000, and the optimal capacity ratio $\eta^* = 1.203$ adds another 175 which is related to backlog costs. This results in the upper horizontal darkgray dotted line at 1175 describing the total expected costs $\mu_i^{(\ell)}(\eta^*)$ allocated to occurrence period i .

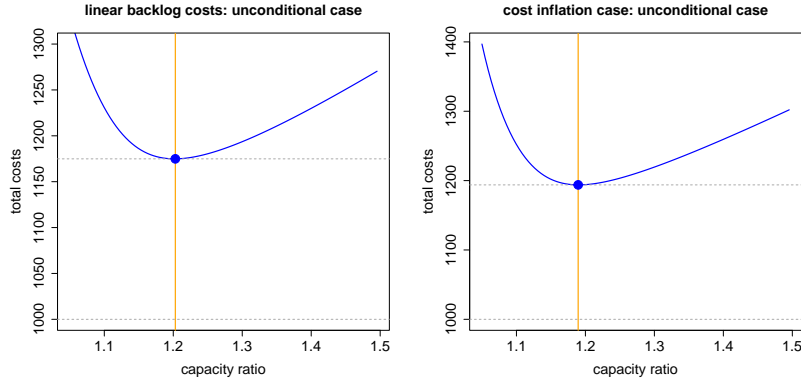


Figure 9: Optimal capacity ratio: (lhs) linear backlog costs case (32) and (rhs) inflating backlog costs case (33); the vertical orange line shows the optimal capacity ratios η^* and the lower horizontal dotted darkgray line the ground-up costs $\kappa_g \mu$.

Figure 9 (rhs) shows the inflating backlog costs case (33), $\eta \mapsto \mu_i^{(\ell)}(\eta)$, where we inflate claims costs by an inflation rate of 5% resulting in $\lambda_b = 1.05$, and the remaining parameters are selected as above. In that case the optimal capacity ratio under our parametrization is $\eta^* = 1.190$.

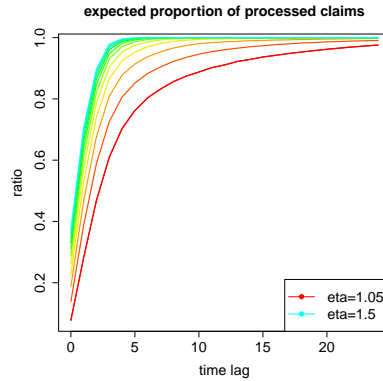


Figure 10: Expected proportion of processed claims for a fixed occurrence period for different capacity ratios $\eta \in \{1.05, \dots, 1.50\}$.

The explicit formula (34) for the expected backlogs $\mathbb{E}[B_{i,j}^{(\eta)}]$ for a fixed occurrence period i for different delays j allows us to compute the expected processing pattern $j \mapsto \mathbb{E}[P_{i,j}^{(\eta)}]$ of a fixed occurrence period, see also Corollary 5.4. Figure 10 shows the cumulative proportion of expected processed claims for different capacity ratios $\eta \in \{1.05, \dots, 1.50\}$. For the higher capacity ratios all claims are likely processed within 5 periods, whereas for lower capacity ratios it may

take up to 24 periods on average. This precisely explains the cost differences implied by the inflation part in the delay-inflated cost case (33).

8.2 Conditional cost optimization

Cost optimization in the conditional case is more involved because we cannot use convenient consequences of stationarity. Here we assume that customers pay a fixed premium and the cost optimization aims to maximize the profit on the business. We consider a going-concern view, assuming that the business continues as planned during a given planning horizon with a constant capacity that we aim to select optimally. We consider all costs in the time window $(\tau, \tau + T]$ without running off the open claims at the end of this time window. For $T \rightarrow \infty$, the results converge to the unconditional case because the impact of the starting configuration at time τ vanishes asymptotically as $T \rightarrow \infty$.

The optimization problem requires us to study $\mathbb{E}[B_{i,\tau-i+k+1}^{(\eta)} | \mathcal{G}_\tau^{(\eta)}]$, where

$$\mathcal{G}_\tau^{(\eta)} = \sigma(R_{i,j}, B_{i,j}^{(\eta)} : i + j \leq \tau, j \geq 0).$$

Hence, we know all ingoing backlogs $(B_{i,j}^{(\eta)})_{i+j \leq \tau}$ into period τ and all reported claims $(R_{i,j})_{i+j \leq \tau}$ in period τ . Having a constant capacity $C_t = c^{(\eta)}$, we therefore also know the aggregated ingoing backlog $B_{\tau+1}^{(\eta)}$ one period later. However, if $B_{\tau+1}^{(\eta)} > 0$, then information $\mathcal{G}_\tau^{(\eta)}$ does not provide its partition to $B_{i,\tau-i+1}^{(\eta)}$, i.e., the individual origin periods. However, we will see that if we aggregate over the occurrence periods i , this split is not necessary.

We consider the cases of linear backlog costs (32). For the moment we drop the costs for the excess capacity $\kappa_c(c^{(\eta)} - \mu)$, since we do not allocate these costs to individual occurrence periods in the going-concern view, we can just add these costs at the end for every period up to the planning horizon T . In the linear cost case we need to study

$$\kappa_g \sum_k \mathbb{E}[R_{i,\tau-i+k+1} | \mathcal{G}_\tau^{(\eta)}] + \kappa_b \sum_k \mathbb{E}[B_{i,\tau-i+k+1}^{(\eta)} | \mathcal{G}_\tau^{(\eta)}], \quad (35)$$

the summation index k is going to be discussed below.

For the subsequent considerations one should have the following matrix in mind. The rows in this matrix reflect the different occurrence periods from $\tau - J$ (first row) to $\tau + T$ in (last row). These are all occurrence periods that contribute to payments in the time window $(\tau, \tau + T]$. Thus, we have $T + J + 1$ occurrence periods in this matrix. The columns in this matrix correspond to calendar periods. The first column reflects the reported claims in calendar period τ , the second column the expected number of reported claims in calendar period $\tau + 1$, and the last column the expected number of reported claims in calendar period

$\tau + T$. Thus, this matrix has $T + 1$ columns.

$$M := \begin{pmatrix} R_{\tau-J,J} & 0 & 0 & 0 & \cdots & 0 & 0 \\ R_{\tau-J+1,J-1} & \mu_J & 0 & 0 & \cdots & 0 & 0 \\ R_{\tau-J+2,J-2} & \mu_{J-1} & \mu_J & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ R_{\tau,0} & \mu_1 & \mu_2 & \mu_3 & \cdots & 0 & 0 \\ 0 & \mu_0 & \mu_1 & \mu_2 & \cdots & 0 & 0 \\ 0 & 0 & \mu_0 & \mu_1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \mu_0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \mu_0 & \mu_1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \mu_0 \end{pmatrix} \quad (36)$$

One should notice that calendar period τ corresponding to the first column aggregates to the total number of reported claims R_τ in calendar period τ , all other columns aggregate to μ being the expected number of claims of futures periods in this stationary model.

The reported claims terms in (35) are comparably simple because future reportings are independent of $\mathcal{G}_\tau^{(\eta)}$. With finite planning horizon we have a total number of expected reported claims, this is the sum over columns 2 to $T + 1$ in matrix (36),

$$\sum_{k=0}^{T-1} \sum_{i=\tau+k+1-J}^{\tau+k+1} \mathbb{E}[R_{i,\tau-i+k+1} | \mathcal{G}_\tau^{(\eta)}] = \sum_{c=2}^{T+1} \sum_{r=c}^{c+J} M_{r,c} = T\mu.$$

Note that occurrence periods $i \leq \tau - J$ are fully reported at time τ , therefore, they do not contribute to the above sum. We add the planning horizon T and we discard all costs after this time window.

Next we focus on the backlogs in (35). For this we come back to the three terms (28)-(30) which need to be summed over k for a fixed occurrence period i . We start with the term (30). This requires $i \leq \tau$, otherwise this occurrence period cannot have any ingoing backlog. Their total contribution across all occurrence periods i is then

$$\sum_{i \leq \tau} B_{i,\tau-i}^{(\eta)} \sum_{k=0}^{T-1} G_\tau^{(\eta)} h_k(B_{\tau+1}^{(\eta)}, 0; \eta) = B_\tau^{(\eta)} \sum_{k=0}^{T-1} G_\tau^{(\eta)} h_k(B_{\tau+1}^{(\eta)}, 0; \eta),$$

we add the planning horizon T and we discard all costs after this time window.

Next, we focus on the term (29) newly reported claims in period τ . This requires occurrence periods $i \leq \tau \leq i + J$, thus, we have total expected backlogs from newly reported claims, this is the sum over the first column in matrix (36),

$$\sum_{i=\tau-J}^{\tau} \frac{R_{i,\tau-i}}{R_\tau} \sum_{k=0}^{T-1} F_\tau^{(\eta)} h_k(B_{\tau+1}^{(\eta)}, 0; \eta) = \sum_{k=0}^{T-1} F_\tau^{(\eta)} h_k(B_{\tau+1}^{(\eta)}, 0; \eta),$$

this is again truncated at the planning horizon T .

Finally, we focus on the term (28) for future reportings for occurrence periods $i > \tau - J$. To verify the following computation one should again note that we aggregate over the columns 2 to $T + 1$ in matrix (36), and each of these columns belongs to a different delay $m \in \{1, \dots, T\}$,

$$\begin{aligned}
& \sum_{i=\tau-J+1}^{\tau+T} \sum_{k=1}^T \sum_{m=1}^k \mathbb{1}_{\{i \leq \tau+m \leq i+J\}} \frac{\mu_{\tau-i+m}}{\mu} h_{k-m} \left(B_{\tau+1}^{(\eta)}, m; \eta \right) \\
&= \sum_{m=1}^T \sum_{i=\tau-J+1}^{\tau+T} \mathbb{1}_{\{\tau-J+m \leq i \leq \tau+m\}} \frac{\mu_{\tau+m-i}}{\mu} \sum_{k=m}^T h_{k-m} \left(B_{\tau+1}^{(\eta)}, m; \eta \right) \\
&= \sum_{m=1}^T \sum_{k=m}^T h_{k-m} \left(B_{\tau+1}^{(\eta)}, m; \eta \right) = \sum_{m=1}^T \sum_{k=0}^{T-m} h_k \left(B_{\tau+1}^{(\eta)}, m; \eta \right).
\end{aligned}$$

Collecting all terms and aggregating over all occurrence periods $i \in \{\tau - J, \dots, T\}$ that contribute to the costs in the time window $(\tau, \tau + T]$, see (36), the linear costs case (35) requires studying the aggregate costs

$$\begin{aligned}
T\mu^{(\ell)}(\eta, T, \mathcal{G}_\tau^{(\eta)}) &:= \kappa_g T \mu + \kappa_b B_\tau^{(\eta)} \sum_{k=0}^{T-1} G_\tau^{(\eta)} h_k \left(B_{\tau+1}^{(\eta)}, 0; \eta \right) \\
&+ \kappa_b \sum_{k=0}^{T-1} F_\tau^{(\eta)} h_k \left(B_{\tau+1}^{(\eta)}, 0; \eta \right) \\
&+ \kappa_b \sum_{m=1}^T \sum_{k=0}^{T-m} h_k \left(B_{\tau+1}^{(\eta)}, m; \eta \right) + \kappa_c T \left(c^{(\eta)} - \mu \right).
\end{aligned} \tag{37}$$

Thus, we have a fixed past history $\mathcal{G}_\tau^{(\eta)}$, we have a fixed planning horizon $T \geq 1$, and we try to minimize the costs in time window $(\tau, \tau + T]$ in the cost capacity ratio η , for given cost parameters $\kappa_g, \kappa_b, \kappa_c > 0$. To keep things simple, we assume that we start from a zero backlog $B_\tau^{(\eta)} = 0$ at time τ . This allows us to drop the backlog term in (37), and we receive a next backlog at time $\tau + 1$

$$B_{\tau+1}^{(\eta)} = \max(R_\tau - C^{(\eta)}, 0) = \max \left(\sum_{i \leq \tau} R_{i, \tau-i} - C^{(\eta)}, 0 \right). \tag{38}$$

Thus, the zero initial backlog case $B_\tau^{(\eta)} = 0$ results in studying

$$\begin{aligned}
T\mu^{(\ell)}(\eta, T, \mathcal{G}_\tau^{(\eta)}) &= T \left(\kappa_c (c^{(\eta)} - \mu) + \kappa_g \mu \right) + \kappa_b \sum_{k=0}^{T-1} F_\tau^{(\eta)} h_k \left(B_{\tau+1}^{(\eta)}, 0; \eta \right) \\
&+ \kappa_b \sum_{m=1}^T \sum_{k=0}^{T-m} h_k \left(B_{\tau+1}^{(\eta)}, m; \eta \right).
\end{aligned}$$

This is fairly simple now. It requires that starting from a zero backlog $B_\tau^{(\eta)} = 0$ at time τ , we need to simulate the number of reported claims R_τ in period τ , from which we can compute the new backlog $B_{\tau+1}^{(\eta)}$ at time $\tau + 1$ as well as $F_\tau^{(\eta)}$. In the following analysis we have $R_\tau = 1310$ from which we can compute the next backlog (38) for the different capacity ratios $\eta > 1$.

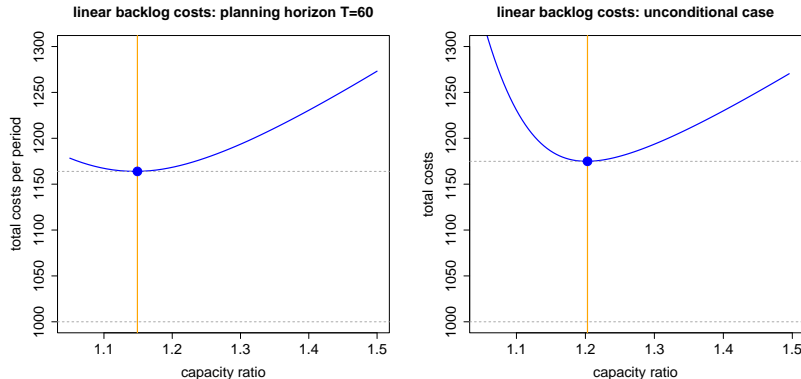


Figure 11: Optimal capacity ratios in the linear backlog costs case, conditional version, with a zero starting backlog $B_\tau^{(\eta)} = 0$ at time τ showing: (lhs) planning horizon $T = 60$, (rhs) unconditional case taken from Figure 9 (lhs), this corresponds to $T = \infty$.

Figure 11 shows the resulting optimal capacity ratios in the conditional case if we start with a zero backlog $B_\tau^{(\eta)} = 0$ at time τ for planning horizon $T = 60$, and it is compared to the unconditional case given in Figure 9 (lhs) using the same cost parameters κ_g , κ_b and κ_c ; note that the unconditional case corresponds to the infinite planning horizon.

planning horizon T	optimal η^*	costs $\mu^{(\ell)}(\eta^*, T, \mathcal{G}_\tau^{(\eta^*)})$
36	1.068	1152
60	1.149	1164
120	1.176	1172
∞	1.203	1175

Table 1: Optimal capacity ratios η^* for different planning horizons T and result total costs linear backlog cost case.

Table 1 gives the numbers for planning horizons including those in Figure 11. With a planning horizon of 120 (monthly) periods (or 10 years) we are rather close to the unconditional case, having average optimal costs per period of 1172. For shorter planning horizons these costs are lower, this is because we start with a zero backlog $B_\tau^{(\eta)} = 0$ at time τ , thus, this is a more favorable starting position than the average over the stationary limit distribution (unconditional

case). This results for a planning horizon of $T = 36$ (monthly) periods (or 3 years) to optimal average costs of 1152.

Similar results with decreasing instead of increasing average costs are obtained if we start from a large initial backlog $B_\tau^{(\eta)}$ at time τ . This can be interpreted as a situation where the backlog has gone out of control, e.g., due to a catastrophic claims event, and the management is concerned about clearing the backlog at minimal mid-term costs. We refrain from giving an explicit numerical example.

9 Summary

We formalized the question of choosing optimal processing capacities for claims handling. On the one hand, the claims handling capacity needs to be limited because any insurance company has only finite financial resources available. On the other hand, the capacity should be sufficiently large because long processing delays (and large backlogs) also generate various costs. We studied this trade-off aiming at minimizing claims and claims processing costs. This problem has several features from queueing theory, but there are also some significant differences because claims are labeled by occurrence periods and arising expenses need to be allocated to occurrence periods to have a consistent and appropriate cost analysis of an insurance portfolio.

We formalized these questions and we solved a variant of this optimal cost and capacity problem. This variant describes a specific mechanism to work off a backlog, and it considers a specific super-imposed cost inflation factor for late claims processing. In this regard, there are many alternative ways to model these backlog cost items. Our choice is a realistic one that is still fairly well tractable, and the final intractable step was solved by a recurrent neural network approximation. This paper appears to be the first that considers this claims processing costs problem. Alternative ways to model delay-adjusted costs and other consequences of backlogs due to capacity constraints with shared capacity could be fruitful to explore. We invite interested scholars to contribute to this interesting problem.

Acknowledgement. Parts of this research was carried out while Mario Wüthrich was a KAW guest professor at Stockholm University. Filip Lindskog acknowledges financial support from the Swedish Research Council, Project 2020-05065.

References

- [1] Albrecher, H., Denuit, M., Trufin, J. (2011). Ruin problems under IBNR dynamics. *Applied Stochastic Models in Business and Industry* 27(6), 619-632.

- [2] Asmussen, S. (2003). *Applied Probability and Queues*, second edn. Springer, New York.
- [3] Billingsley, P. (1995). *Probability and Measure*, third edn. Wiley, New York.
- [4] Boogaert, P., Haezendonck, J. (1989). Delay in claim settlement. *Insurance: Mathematics and Economics* 8, 321-330.
- [5] Buchwalder, M., Bühlmann, H., Merz, M., Wüthrich, M.V. (2006). Estimation of unallocated loss adjustment expenses. *Bulletin of the Swiss Association of Actuaries* 2006(1), 43-53.
- [6] Chen, Y., Whitt, W. (2020). Algorithms for the upper bound mean waiting time in the GI/GI/1 queue. *Queueing Systems* 94, 327-356.
- [7] Daley, D.J. (1977). Inequalities for moments of tails of random variables, with queueing applications. *Zeitschrift für Wahrscheinlichkeitstheorie Verw. Gebiete* 41, 139-143.
- [8] Harel, A. (1990) Convexity results for single-serves queues and multiserver queues with constant service times. *Journal of Applied Probability* 27(2), 465-468.
- [9] Huynh, M., Landriault, D., Shi, T., Willmot, G.E. (2015). On a risk model with claim investigation. *Insurance: Mathematics and Economics* 65, 37-45.
- [10] Janssen, A.J.E.M., Van Leeuwen, J.S.H. (2005). Relaxation time for the discrete D/G/1 queue. *Queueing Systems* 50, 53-80.
- [11] Janssen, A.J.E.M., Van Leeuwen, J.S.H. (2018). Spitzer's identity for discrete random walks. *Operations Research Letters* 46, 168-172.
- [12] Kingman, J.F.C. (1962). Inequalities for the queue GI/G/1. *Biometrika* 49(3/4), 315-324.
- [13] Maglaras, C., Zeevi, A. (2003). Pricing and capacity sizing for systems with shared resources: approximate solutions and scaling relations. *Management Science* 49(8), 1018-1038.
- [14] Sato, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Studies in Advanced Mathematics 68, Cambridge University Press.
- [15] Stadje, W. (1997). A new approach to the Lindley recursion. *Statistics & Probability Letters* 31, 169-175.
- [16] Waters, H.R., Papatriandafylou, A. (1985). Ruin probabilities allowing for delay in claims settlement. *Insurance: Mathematics and Economics* 4, 113-122.

A Figures for neural network approximations

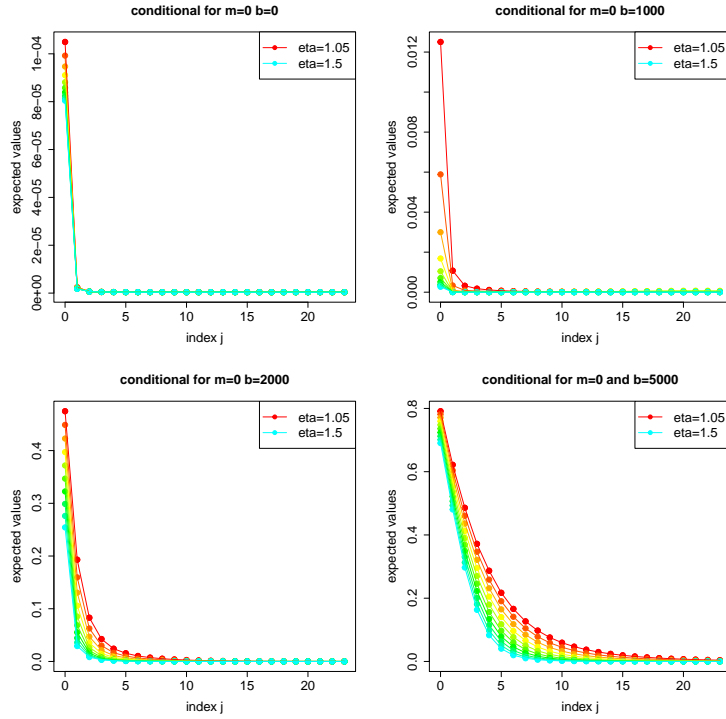


Figure 12: RNN approximations of the conditional means $h_j(b, 0; \eta)$, $j \geq 1$, for capacity ratios $\eta \in \{1.05, 1.10, \dots, 1.50\}$ (in different colors), $b \in \{0, 1000, 2000, 5000\}$ (different plots) and $m = 0$.

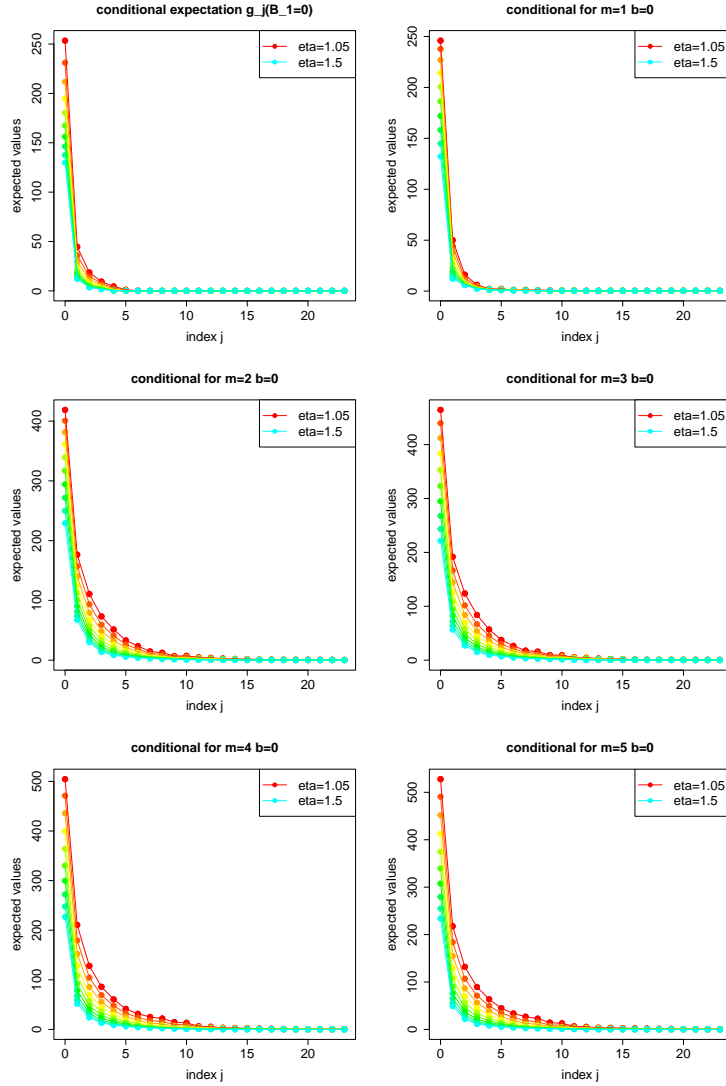


Figure 13: RNN approximations of the conditional means $h_j(b, m; \eta)$, $j \geq 0$, for capacity ratios $\eta \in \{1.05, 1.10, \dots, 1.50\}$ (in different colors), $b = 0$ and $m \in \{1, \dots, 5\}$ (different plots), top-left is taken from Figure 6.

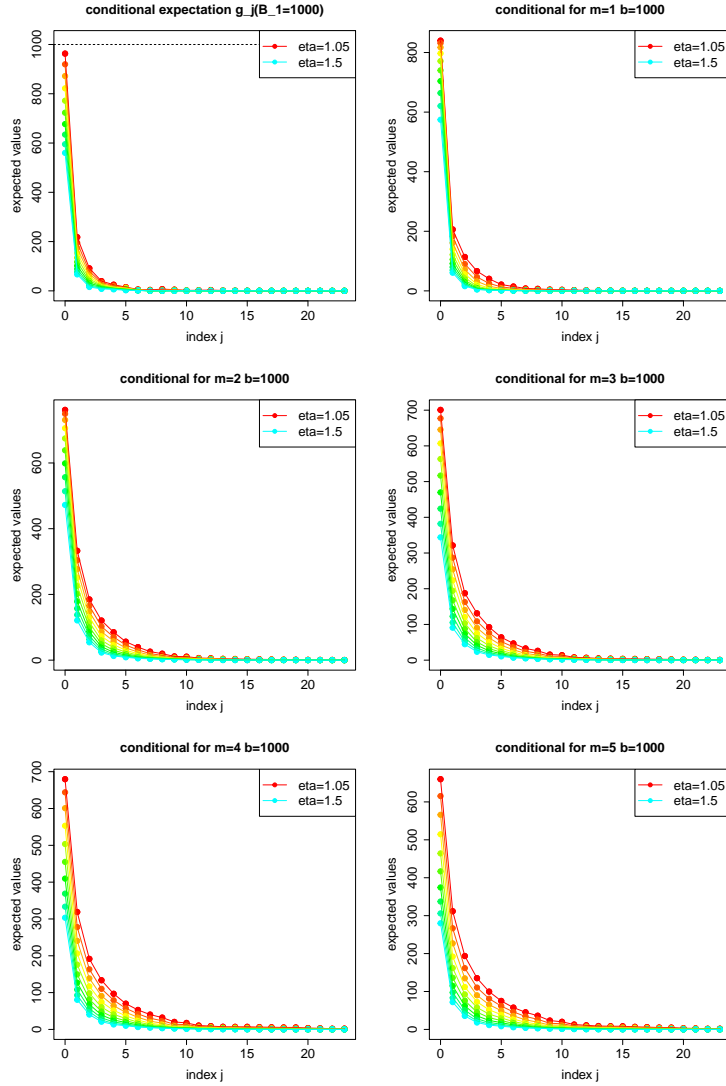


Figure 14: RNN approximations of the conditional means $h_j(b, m; \eta)$, $j \geq 0$, for capacity ratios $\eta \in \{1.05, 1.10, \dots, 1.50\}$ (in different colors), $b = 1000$ and $m \in \{1, \dots, 5\}$ (different plots), top-left is taken from Figure 6.

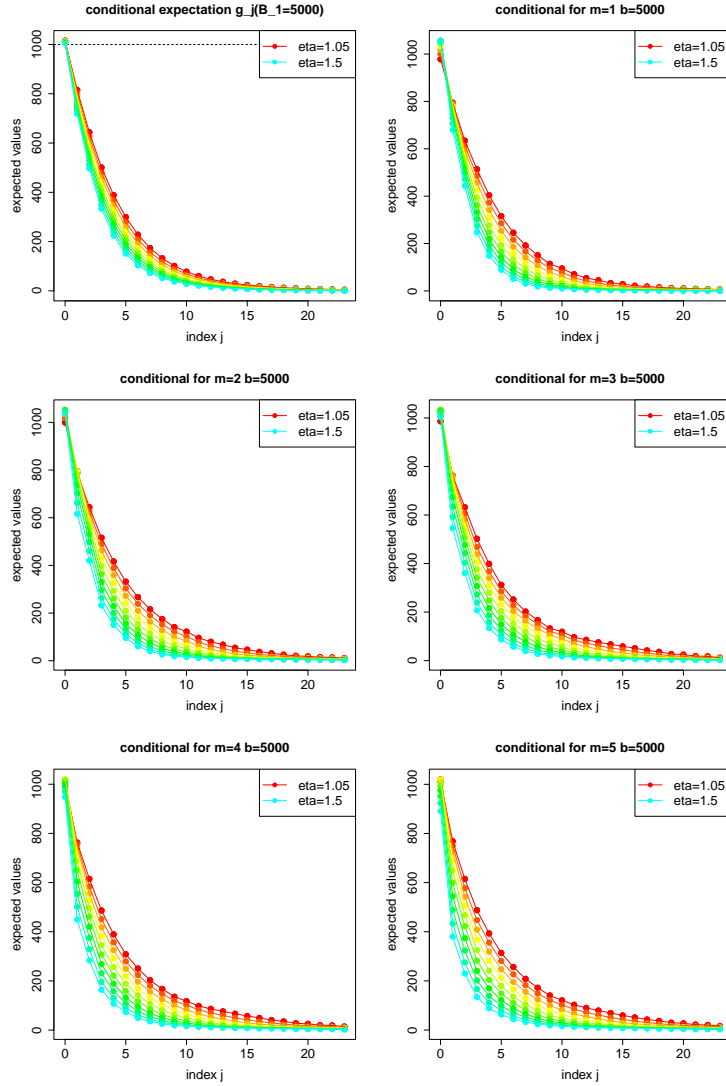


Figure 15: RNN approximations of the conditional means $h_j(b, m; \eta)$, $j \geq 0$, for capacity ratios $\eta \in \{1.05, 1.10, \dots, 1.50\}$ (in different colors), $b = 5000$ and $m \in \{1, \dots, 5\}$ (different plots), top-left is taken from Figure 6.