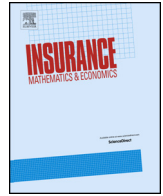




ELSEVIER

Contents lists available at ScienceDirect

## Insurance: Mathematics and Economics

journal homepage: [www.elsevier.com/locate/ime](http://www.elsevier.com/locate/ime)The Cramér–Lundberg model with a fluctuating number of clients <sup>☆</sup>Peter Braunsteins <sup>a,b</sup>, Michel Mandjes <sup>a,c,d,\*</sup><sup>a</sup> Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands<sup>b</sup> Mathematical Institute, P.O. Box 9512, 2300 RA Leiden, the Netherlands<sup>c</sup> EURANDOM, Eindhoven University of Technology, Eindhoven, the Netherlands<sup>d</sup> Amsterdam Business School, Faculty of Economics and Business, University of Amsterdam, Amsterdam, the Netherlands

## ARTICLE INFO

## Article history:

Received August 2022

Received in revised form May 2023

Accepted 24 May 2023

Available online 30 May 2023

## JEL classification:

6 C

149 P

## Keywords:

Cramér–Lundberg

Large deviations

Ruin probability

Exponential tightness

## ABSTRACT

This paper considers the Cramér–Lundberg model, with the additional feature that the number of clients can fluctuate over time. Clients arrive according to a Poisson process, where the times they spend in the system form a sequence of independent and identically distributed non-negative random variables. While in the system, every client generates claims and pays premiums. In order to describe the model's rare-event behaviour, we establish a sample-path large-deviation principle. This describes the joint rare-event behaviour of the reserve-level process and the client-population size process. The large-deviation principle can be used to determine the decay rate of the time-dependent ruin probability as well as the most likely path to ruin. Our results allow us to determine whether the chance of ruin is greater with more or with fewer clients and, more generally, to determine to what extent a large deviation in the reserve-level process can be attributed to an unusual outcome of the client-population size process.

© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Cramér–Lundberg (CL) model (Cramér, 1930; Lundberg, 1903, 1926) plays a pivotal role in ruin theory. It is a stochastic process that represents the evolution of an insurance firm's reserve level (also referred to as surplus-level process). The primary goal is to evaluate the *ruin probability* for a given initial surplus  $u$ , i.e., the probability that the reserve-level process drops below 0. In the most basic variant of the CL model, claims are independent and identically distributed (iid) non-negative random quantities that arrive according to a Poisson process (with rate  $\nu > 0$ ), while premiums are earned at a deterministic linear rate  $r > 0$ . For this base model a broad range of results has been obtained, most notably a characterisation of the ruin probability through its Laplace transform. In addition, relying on elements from *large-deviations theory*, the asymptotics of the ruin probability were identified for large values of the initial surplus  $u$ , an important observation in this context being that with overwhelming probability the path to ruin is by approximation linear (under the proviso that the claims are light-tailed). For more background on these results, and an account of the area of ruin theory in general, we refer to e.g. Asmussen and Albrecher (2010); Embrechts et al. (1997); Kyprianou (2013); Rolski et al. (2009).

The CL model described above is admittedly a gross simplification of reality, in that various features that play a role in practice are not incorporated. This realization led to a stream of results that in various directions generalize the classical setup. Without attempting to provide an exhaustive overview, we now briefly mention a few of the main strands of research. Arguably the most important extension concerns the time-dependent ruin probability, i.e., the probability that the reserve level becomes negative before a given point in time. We refer to Asmussen and Albrecher (2010, Ch. V) for an overview of results in this area; notably, under the large-deviations scaling (with light-tailed claims) the most likely path to ruin is still linear. Steps have also been taken to generalise the claim arrival process, which is traditionally of compound Poisson type. In Gerber (1970) a diffusion term is added, and (more generally) in Dębicki and Mandjes (2015); Kyprianou (2006) the reserve level evolves as a Lévy process. In e.g. Stabile and Torrisi (2010) the arrival process is assumed to be of Hawkes type. Other extensions include variants in which the insurance firm's interest income is incorporated; see for instance (Albrecher

<sup>☆</sup> Both authors' research has been funded by the NWO Gravitation project NETWORKS, grant number 024.002.003.

\* Corresponding author.

E-mail addresses: [pbraunsteins@gmail.com](mailto:pbraunsteins@gmail.com) (P. Braunsteins), [m.r.h.mandjes@uva.nl](mailto:m.r.h.mandjes@uva.nl) (M. Mandjes).

et al., 2013; Boxma and Mandjes, 2021), and the textbook treatment in Asmussen and Albrecher (2010, Ch. VIII). We finally mention the branch of the literature in which the reserve process is modulated by a background process; see e.g. the Markov-modulated framework in Asmussen and Albrecher (2010, Ch. VII) and the mixing model in Constantinescu et al. (2020).

In the present paper we consider another extension of the CL model, namely a model in which the insurance firm has a stochastically fluctuating number of clients. One could view the standard CL model as a setup in which the number of clients is fixed, while in practice, so as to properly assess the ruin probability, one should evidently take into account variations in the client population size. We model the client-level fluctuations by letting clients arrive according to a Poisson process, where the times they spend in the system (as a client of the insurance firm, that is) form a sequence of iid non-negative random variables; while in the system, each client generates iid claims at Poisson instants, and pays premiums at a rate  $r$ .

In the CL model with a fluctuating number of clients, we wish to assess the time-dependent ruin probability, given the insurance firm’s initial surplus. We do so in an asymptotic context, corresponding to the (realistic) situation that the insurance firm’s client base is consistently large. Concretely, we let the (Poissonian) client arrival rate be  $n\lambda$  and the initial surplus be  $nu$  for some  $u > 0$ , where  $n$  is a scaling parameter that we let grow large. In this limiting setting we derive a sample-path large-deviation principle (LDP). This sample-path LDP is bivariate, in that it jointly describes the reserve-level process and the client-population-size process. We use it not only to evaluate the logarithmic decay rate of the time-dependent ruin probability, but also to investigate two questions about the most likely path to ruin: (i) is the chance of ruin greater when the client population is higher or lower than expected?; (ii) to what extent can a large deviation in the reserve level process be attributed to an unusual outcome of the client population process?

At a technical level, the crucial difference with the conventional CL model, where the number of clients is fixed, is that when we allow the number of clients to fluctuate, the increments of the reserve level process are no longer independent. Traditionally, sample-path large deviations mainly focus on settings with independent increments. Results in this area essentially go back to an early paper by Varadhan (1966); see also the contributions in Borovkov (1965); Mogulskii (1974, 1976). Indeed, the sample path LDP for the standard CL model is implied by the classical result of Mogulskii (Dembo and Zeitouni, 1998, Thm. 5.1.2). Models in which there is correlation in the increment process are substantially harder to deal with, but often offer richer behaviour. For example, in the CL model with a fluctuating number of clients the most likely path to ruin is no longer linear. For work on sample-path large deviations for processes with dependent increments we refer to (the generalised version of) Schilder’s theorem for Gaussian processes, which was established in Azencott (1978); Bahadur and Zabell (1979); see also the textbook treatment in Deuschel and Stroock (1989). This type of result has been applied extensively in the operations research domain, addressing various rare-event related problems concerning Gaussian storage systems (Arendarczyk et al., 2014; Mandjes et al., 2006, 2009; Mandjes and van Uitert, 2005; Mannersalo and Norros, 2002). We also point to sample-path LDPs for specific queueing models which can be found in e.g. Blanchet et al. (2014); Shwartz and Weiss (1995); Wischik (2001).

We prove the LDP for our variant of the CL model by first establishing an LDP that corresponds to a single point in time, then extending this to an LDP for multiple points in time, before finally establishing the full sample-path LDP. In this approach, the first two steps rely on a fundamental observation: even though its increments are not independent, it is possible to decompose the process into independent components, thus allowing arguments based on sums of independent random variables to be applied. The main technical hurdle lies in the final step: upgrading the finite-dimensional LDP to a sample-path LDP. This amounts to verifying one of the equivalent exponentially tightness characterisations as provided by Feng and Kurtz (2006, Thm. 4.1). We point out that since the number of clients fluctuates autonomously (i.e., it is not affected by the reserve-level process), the structure of the LDP resembles the decompositions found in Huang et al. (2016); Liptser (1996).

This paper is organised as follows. In Section 2 we provide a detailed model description of our CL model with a fluctuating client population. In Section 3 we present our main results. These cover finite-dimensional LDPs as well as the full sample-path LDP. In addition, we present results that shed light on the most likely path to ruin, including experimental insight into the most likely cause of ruin. Proofs are provided in Section 4: first we focus on establishing finite-dimensional LDPs, and then extend these to the full sample-path LDP by relying on a tightness argument.

## 2. Model

In this section we introduce the CL model with a fluctuating client population. In this model description, we distinguish between the dynamics of the population size, and the dynamics corresponding to each individual client in the system.

*Client-population-size dynamics.* Clients arrive to the system according to a Poisson process with rate  $n\lambda$ . Here  $\lambda$  is a positive parameter, and  $n$  is a scaling parameter that we let grow large. The clients stay in the system for independent and identically distributed (iid) amounts of time, in the sequel referred to as the clients’ *sojourn times*. For convenience, in our analysis we let the sojourn times have density  $h(\cdot)$ , but our arguments hold more generally (in particular allowing for both continuous and discrete sojourn-times distributions). In queueing-theoretic terminology, the number of clients simultaneously present follows the dynamics of a so-called  $M/G/\infty$  system.

At time 0, the number of clients already present is  $nf_0$  for some  $f_0 \geq 0$ . These  $nf_0$  clients have remaining sojourn times that are iid with density  $h^\circ(\cdot)$ . In this respect a natural choice is to let the remaining sojourn times have the well-known *excess lifetime distribution*, i.e., for  $t \geq 0$ ,

$$h^\circ(t) = \frac{\int_t^\infty h(s) ds}{\int_0^\infty sh(s) ds} ,$$

where in the denominator we recognize the mean duration of a ‘fresh’ sojourn time; it is easily verified that this density integrates to 1.

Recall that for the  $M/G/\infty$  system in equilibrium, the number of clients simultaneously present has a Poisson distribution with mean

$$\lambda \int_0^\infty sh(s) ds .$$

Moreover, their remaining service times are independent and obey the excess lifetime distribution, independently of the number of clients present.

Throughout we impose the mild technical assumption that remaining sojourn times have a uniformly bounded density, i.e., that there exists a constant  $C < \infty$  such that  $h^\circ(t) \leq C$  for all  $t \geq 0$ . Note that this assumption holds if  $h^\circ(\cdot)$  is the excess lifetime distribution (as then we have that  $h^\circ(t)$  is, for any  $t \geq 0$ , majorised by the multiplicative inverse of the mean of a ‘fresh’ sojourn time). A minor technical remark is that, for convenience, the number  $nf_0$  is throughout assumed to be an integer, but in the case it is not integer our analysis can be adapted easily by a straightforward rounding procedure.

Let  $(T_{i,n})_{i \geq 1}$  denote the sequence of iid exponentially distributed random variables describing the clients’ arrival times in the  $n$ -th process and  $(\mathcal{N}_{t,n})_{t \geq 0}$  the corresponding renewal process (i.e., a Poisson process of rate  $n\lambda$ ). Let  $(\tau_i)_{i \geq 1}$  denote the iid sequence of sojourn times and let  $(\tau_i^\circ)_{i \geq 1}$  denote the iid sequence of remaining sojourn times. The number of clients present at  $t$  is then

$$\begin{aligned} F_n(t) &:= nf_0 - \sum_{i=1}^{nf_0} 1\{\tau_i^\circ \leq t\} + \mathcal{N}_{t,n} - \sum_{i=1}^{\mathcal{N}_{t,n}} 1\{T_{i,n} + \tau_i \leq t\} \\ &= \sum_{i=1}^{nf_0} 1\{\tau_i^\circ > t\} + \sum_{i=1}^{\mathcal{N}_{t,n}} 1\{T_{i,n} + \tau_i > t\}. \end{aligned} \tag{1}$$

Notice that  $F_n(t)$  consists of both clients who belonged to the initial  $nf_0$  clients (and have not left yet by time  $t$ ) and clients who arrived in  $(0, t]$  (and are still present at time  $t$ ). We denote the corresponding normalised process by

$$\bar{F}_n(t) := \frac{F_n(t)}{n}.$$

*Client behaviour.* Now that we have introduced the stochastic mechanism that generates the client-population dynamics, we continue by focusing on the behaviour of each individual client while being in the system. During her sojourn time a client pays a constant premium rate of  $r > 0$  per unit of time. Every client generates claims at a Poisson rate  $\nu > 0$  while in the system. The claim sizes form an iid sequence, with the moment generating function (mgf) of an individual claim being denoted by  $\beta(\cdot)$ . Throughout we assume that we are in the light-tailed setting, in that  $\beta(\theta)$  is finite for  $\theta$  in an open neighbourhood of the origin. The *net aggregate claim process* represents the total claimed amount (by the entire population, that is) decreased by the premiums received by the insurance firm. Let  $(\mathcal{M}_{t,i}^\circ)_{t \geq 0}$  and  $(\mathcal{M}_{t,i})_{t \geq 0}$  denote independent sequences of Poisson processes of rate  $\nu$ , describing the number of claims corresponding to the initially present and arriving customers respectively, and let  $(Z_{k,i}^\circ)$  and  $(Z_{k,i})$  denote sequences of iid random variables describing the  $k$ -th claim by the initially present and arriving customers respectively. The net aggregate claim process at time  $t \geq 0$  (with  $G_n(0) = 0$ ) is then

$$\begin{aligned} G_n(t) &:= \sum_{i=1}^{nf_0} \left( -r(t \wedge \tau_i^\circ) + \sum_{k=1}^{\mathcal{M}_{t \wedge \tau_i^\circ, i}^\circ} Z_{k,i}^\circ \right) \\ &\quad + \sum_{i=1}^{\mathcal{N}_{t,n}} \left( -r(0 \vee [(t - T_{i,n}) \wedge \tau_i]) + \sum_{k=1}^{\mathcal{M}_{(t - T_{i,n}) \wedge \tau_i, i}} Z_{k,i} \right). \end{aligned} \tag{2}$$

We denote the corresponding normalised process by

$$\bar{G}_n(t) := \frac{G_n(t)}{n}.$$

Our goal is to produce a probabilistic description of the object  $(F_n(\cdot), G_n(\cdot))$  that allows us to identify the logarithmic decay rate (as  $n$  grows large) of the time-dependent ruin probability

$$p_n(u, T) := \mathbb{P}(\exists t \in [0, T] : \bar{G}_n(t) \geq u), \tag{3}$$

given that  $\bar{F}_n(0) = f_0$  and  $\bar{G}_n(0) = 0$ .

### 3. Main results

#### 3.1. Large-deviation principles

Our main result is the sample-path LDP of the bivariate process  $(F_n(\cdot), G_n(\cdot))$ , to be presented in Theorem 1. We establish this LDP by first proving more basic, finite-dimensional LDPs, which we then upgrade to the full sample-path LDP through a tightness argument. Concretely, we first discuss a one-point LDP (pertaining to a single point in time, that is), then extend this to a finite-point LDP (pertaining to finitely many time epochs), and then finally to a sample-path LDP. In this section we state these results, and provide the main ideas behind the proofs (which are given in detail in Section 4).

It is noted that the process  $(F_n(\cdot), G_n(\cdot))$  is not necessarily Markovian – or, more precisely: only when the clients’ sojourn times are exponentially distributed,  $(F_n(\cdot), G_n(\cdot))$  is a Markov chain. Importantly, however, we can still use arguments that are based on sums of independent random variables. Two crucial observations in this context are:

- The process  $(F_n(\cdot), G_n(\cdot))$ , as defined via (1) and (2), can be decomposed as the sum of two independent components: one related to the contribution of the  $nf_0$  clients who were already present at time 0, which we denote by  $(F_n^-(\cdot), G_n^-(\cdot))$ , and one related to the  $\mathcal{N}_{t,n}$  clients who enter in the interval  $(0, t]$ , which we denote by  $(F_n^+(\cdot), G_n^+(\cdot))$ . To be precise,  $F_n^-(t)$  is the number of clients who were present at time 0 who are still present at time  $t$ , and  $G_n^-(t)$  is the net aggregate claim volume up to time  $t$  which was generated by the clients who were present at time 0. The process  $(F_n^+(\cdot), G_n^+(\cdot))$  is defined similarly, but now corresponding to clients who were not present at time 0. As a consequence,  $F_n^+(t) = F_n(t) - F_n^-(t)$  and  $G_n^+(t) = G_n(t) - G_n^-(t)$ .
- By a direct application of known thinning and superposition properties of Poisson processes, both  $(F_n^-(\cdot), G_n^-(\cdot))$  and  $(F_n^+(\cdot), G_n^+(\cdot))$  can be interpreted as sums of iid processes, each of them distributed as some  $(F^-(\cdot), G^-(\cdot))$  and  $(F^+(\cdot), G^+(\cdot))$ , respectively. More precisely,  $(F_n^-(\cdot), G_n^-(\cdot))$  can be represented by the sum of  $nf_0$  iid copies of  $(F^-(\cdot), G^-(\cdot))$ , where each copy corresponds to the contribution of a single client who is present at time 0. Similarly,  $(F_n^+(\cdot), G_n^+(\cdot))$  can be seen as the sum of  $n$  iid copies of  $(F^+(\cdot), G^+(\cdot))$ , where each copy corresponds to the contribution of a stream of clients that arrive according to a Poisson process with rate  $\lambda$ .

### 3.1.1. One-point LDP

We start by deriving a large-deviation principle for the random vector  $(\bar{F}_n(t), \bar{G}_n(t))$  for a given time point  $t > 0$ . As an immediate consequence of the observations above, we can represent  $(\bar{F}_n(t), \bar{G}_n(t))$  as a sum of  $n$  iid random vectors. We can thus apply Cramér’s theorem (Dembo and Zeitouni, 1998, Section 2.2.2), so as to obtain a large-deviation principle whose rate function is given by the Legendre transform, for  $(f, g) \in \mathbb{R}_+ \times \mathbb{R}$ ,

$$I_t(f, g) = \sup_{\omega, \theta \in \mathbb{R}} \{ \omega f + \theta g - f_0 \log M_t^-(\omega, \theta) - \log M_t^+(\omega, \theta) \}, \tag{4}$$

where  $M_t^i(\omega, \theta) = \mathbb{E} \exp(\omega F^i(t) + \theta G^i(t))$  for  $i \in \{-, +\}$ . Cramér’s theorem concretely entails that for a set  $B \subset \mathbb{R}_+ \times \mathbb{R}$  we have that

$$\begin{aligned} - \inf_{(f, g) \in B^\circ} I_t(f, g) &\leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}((\bar{F}_n(t), \bar{G}_n(t)) \in B) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}((\bar{F}_n(t), \bar{G}_n(t)) \in B) \leq - \inf_{(f, g) \in \bar{B}} I_t(f, g), \end{aligned} \tag{5}$$

where  $B^\circ$  is the interior of the set  $B$  and  $\bar{B}$  its closure. It provides an informal justification for the frequently used approximation

$$\mathbb{P}((\bar{F}_n(t), \bar{G}_n(t)) \in B) \approx \exp\left(-n \inf_{(f, g) \in B} I_t(f, g)\right).$$

The next step is to compute the mgfs  $M_t^+(\omega, \theta)$  and  $M_t^-(\omega, \theta)$ . To this end, observe that the net claim process of an individual client (while in the system) is a Lévy process (Dębicki and Mandjes, 2015), viz. a compound Poisson process with drift, say  $Z(\cdot)$ . It is directly verified that the mgf of  $Z(t)$  can be written as  $(\varphi(\theta))^t$ , where

$$\varphi(\theta) = \mathbb{E} \exp(\theta Z(1)) = \exp(-r\theta + v(\beta(\theta) - 1)). \tag{6}$$

To compute  $M_t^-(\omega, \theta)$ , let  $\tau^\circ$  be the random variable corresponding to a typical residual sojourn-time duration of a client who is present at time 0. Conditioning on the time this client leaves, we readily obtain

$$M_t^-(\omega, \theta) = \mathbb{E} e^{\theta Z(\tau^\circ \wedge t)} e^{\omega 1_{\{\tau^\circ > t\}}} = \int_0^t h^\circ(s) (\varphi(\theta))^s ds + (\varphi(\theta))^t e^{\omega} \int_t^\infty h^\circ(s) ds. \tag{7}$$

The next goal is to compute  $M_t^+(\omega, \theta)$ . To this end, we rely on the property that the number of clients that arrive in the interval  $(0, t]$  is Poisson with parameter  $\lambda t$ . In addition, conditional on the number of arrivals, the arrival times can be seen as order statistics of a sequence of iid uniformly distributed random variables, see for example (Ross, 2002, p. 303). We thus find, with  $U$  being uniformly distributed on  $[0, 1]$  and  $\tau$  the random variable corresponding to a typical duration of the time a client spends in the system,

$$\begin{aligned} M_t^+(\omega, \theta) &= \sum_{k=0}^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} \left( \mathbb{E} e^{\theta Z(\tau \wedge t(1-U))} e^{\omega 1_{\{\tau > t(1-U)\}}} \right)^k \\ &= \sum_{k=0}^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} \left( \int_0^t \frac{1}{t} \left( \int_0^{t-s} h(r) (\varphi(\theta))^r dr + (\varphi(\theta))^{t-s} e^{\omega} \int_{t-s}^\infty h(r) dr \right) ds \right)^k, \end{aligned}$$

which simplifies to

$$\exp\left(\lambda \left( \int_0^t \left( \int_0^{t-s} h(r) (\varphi(\theta))^r dr + (\varphi(\theta))^{t-s} e^{\omega} \int_{t-s}^\infty h(r) dr \right) ds - 1 \right)\right). \tag{8}$$

Upon combining the above, we have thus established the following result.

**Proposition 1.** *The pair  $(\bar{F}_n(t), \bar{G}_n(t))$  satisfies the LDP with rate  $n$  and rate function  $I_t(f, g)$  characterised by (4)–(8).*

### 3.1.2. Multi-point LDP

We proceed by deriving a multi-point LDP, i.e., an LDP for the  $2d$ -dimensional random vector

$$(\bar{F}_n(t_1), \dots, \bar{F}_n(t_d), \bar{G}_n(t_1), \dots, \bar{G}_n(t_d))$$

where  $0 \leq t_1 < t_2 < \dots < t_d$  and  $d \in \mathbb{N}$ . This can be seen as the  $d$ -dimensional counterpart of the LDP above: the 2-dimensional vector  $(\bar{F}_n(t), \bar{G}_n(t))$  has to be replaced by the  $2d$ -dimensional vector  $(\bar{F}_n(t_1), \dots, \bar{F}_n(t_d), \bar{G}_n(t_1), \dots, \bar{G}_n(t_d))$  in (5).

Mimicking the argumentation used in the case  $d = 1$ , we now apply the  $2d$ -variate version of Cramér’s theorem (Dembo and Zeitouni, 1998, Section 2.2.2) to obtain an LDP with rate function, for  $(\mathbf{f}, \mathbf{g}) \in \mathbb{R}_+^d \times \mathbb{R}^d$ ,

$$I_t(\mathbf{f}, \mathbf{g}) = \sup_{\theta, \omega \in \mathbb{R}^d} \left( \sum_{j=1}^d \omega_j f_j + \sum_{j=1}^d \theta_j g_j - f_0 \log M_t^-(\omega, \theta) - \log M_t^+(\omega, \theta) \right),$$

where

$$M_t^i(\omega, \theta) = \mathbb{E} \exp \left( \sum_{j=1}^d \omega_j F^i(t_j) + \sum_{j=1}^d \theta_j G^i(t_j) \right), \tag{9}$$

for  $i \in \{-, +\}$ . Note that, as before, we split the required mgf into one representing the contribution of the clients present at time 0 and another corresponding to the contribution of the clients arriving in  $(0, t]$ , with these two contributions being independent. We can derive the mgfs  $M_t^-(\omega, \theta)$  and  $M_t^+(\omega, \theta)$  by following a similar method to the one used in the one-point case; however, due to the non-Markovian nature of the process, this derivation is relatively involved and is therefore postponed to Section 4.1. We thus establish the following result.

**Proposition 2.** *The vector  $(\bar{F}_n(t_1), \dots, \bar{F}_n(t_d), \bar{G}_n(t_1), \dots, \bar{G}_n(t_d))$  satisfies the LDP with rate  $n$  and rate function  $I_t(\mathbf{f}, \mathbf{g})$ , where  $M_t^-(\omega, \theta)$  and  $M_t^+(\omega, \theta)$  are given in Lemmas 1 and 2, respectively.*

### 3.1.3. Sample-path LDP

The next step is to extend the LDP for finitely many points in time to a full sample-path LDP on  $D(\mathbb{R}^2, [0, T])$ , the space of  $\mathbb{R}^2$ -valued càdlàg functions endowed with the Skorokhod topology, with rate function  $I_{[0, T]}(f, g)$  defined later in (24). Roughly speaking, this is done in two steps: (i) we derive limiting expressions for  $M_t^-(\omega, \theta)$  and  $M_t^+(\omega, \theta)$  as the mesh  $0 = t_1 < t_2 < \dots < t_d = T$  becomes infinitely fine (done in Section 4.2.1); (ii) we prove that the sequence of processes  $(\bar{F}_n(\cdot), \bar{G}_n(\cdot))$  is exponentially tight (done in Section 4.2.2). As it turns out, from a computational perspective it is easier to work with a different expression for the rate function  $I_{[0, T]}(f, g)$ : as pointed out in Section 4.2.3 we can decompose  $I_{[0, T]}(f, g)$  into two parts under the proviso that both  $f$  and  $g$  are absolutely continuous. On the other hand, when  $f$  or  $g$  is not absolutely continuous, then we show (also in Section 4.2.2) that  $I_{[0, T]}(f, g) = \infty$ . Formally, our LDP result is summarized in the following statement.

**Theorem 1.** *The sequence of processes  $(\bar{F}_n(t), \bar{G}_n(t))_{t \geq 0}$  satisfies the LDP on  $D(\mathbb{R}^2, [0, T])$  with rate  $n$  and rate function  $I_{[0, T]}(f, g)$  characterised by (24) and Lemmas 5 and 6.*

## 3.2. Experiments

Evidently, the primary application of Theorem 1 is to evaluate the decay rate of the time-dependent ruin probability  $p_n(u, T)$ , as was defined in (3), in our model with a fluctuating number of clients. In addition, however, it reveals the most likely way in which rare events, such as the insurance firm going bankrupt, occur. In this subsection we apply Theorem 1 to explore in detail two areas of interest, both of them related to the most likely path to bankruptcy.

- (1) What is the most likely path of  $(\bar{F}_n(\cdot), \bar{G}_n(\cdot))$  to bankruptcy at some time  $T$ ? More specifically, is the insurance firm more likely to go bankrupt when there are more clients than usual or fewer clients than usual? We remark that the answer to this question is not *a priori* obvious: more clients mean more revenue, but also a higher risk of large claims. This question will be systematically analyzed in Section 3.2.1.
- (2) What is the primary cause of fluctuations in the reserve level at some time  $T$ ? Specifically, when are these fluctuations primarily due to randomness in the number of clients, and when are they primarily due to randomness in the claims made by these clients? We shed light on this issue in Section 3.2.2.

As argued below, in the context of both questions, a crucial role is played by the probability that the process  $(\bar{F}_n(\cdot), \bar{G}_n(\cdot))$  is in the ‘ruin set’  $\mathcal{R} := [0, \infty) \times B$  at time  $t \geq 0$  with  $B \subset \mathbb{R}$ , i.e., the probability that  $(\bar{F}_n(\cdot), \bar{G}_n(\cdot))$  belongs to

$$\mathcal{H}_t := \{(f, g) : (f(t), g(t)) \in \mathcal{R}\}. \tag{10}$$

In light of Theorem 1, in order to find the logarithmic decay rate of this probability we are to solve the variational problem

$$\varrho(t) := \inf_{(f, g) \in \mathcal{H}_t} I_{[0, t]}(f, g). \tag{11}$$

All numerical results included in this section are obtained using the method that is outlined in Appendix A.

### 3.2.1. Path to bankruptcy

We consider the situation that  $\mathcal{R} = [0, \infty) \times [u, \infty)$ , where  $u$  corresponds to the initial surplus of the insurance firm. This means that, due to Theorem 1, the logarithmic decay rate of the time-dependent ruin probability can be found by solving the following optimisation:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log p_n(u, T) = - \inf_{t \in [0, T]} \varrho(t).$$

We start, however, by studying the probability of *eventual* bankruptcy, i.e., bankruptcy over an infinite horizon (in the literature also frequently referred to as the *all-time ruin probability*). To this end, we consider

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log p_n(u, \infty) = -\varrho^* := - \inf_{t \geq 0} \varrho(t).$$

Let  $t^*$  denote the corresponding optimising time (so that  $\varrho^* = \varrho(t^*)$ ), and  $f^*, g^*$  be the corresponding optimising paths (so that  $\varrho^* = I_{[0, t^*]}(f^*, g^*)$ ). The next proposition reflects the remarkable fact that the probability of eventual bankruptcy is independent of fluctuations in the number of clients. An explanation of this fact is given below. Note that this result holds not only when  $\mathcal{R}$  is of the form  $[0, \infty) \times [u, \infty)$ , but more generally when  $\mathcal{R} = B \times [0, \infty)$  with  $B \subset \mathbb{R}$ .

**Proposition 3.** *If  $\mathcal{R} = B \times [0, \infty)$  with  $B \subset \mathbb{R}$ , then  $\varrho^*$  is independent of the client-level dynamics (i.e.,  $f_0, \lambda, h^\circ(\cdot)$ , and  $h(\cdot)$ ). In addition,  $\varrho^*$  only depends on  $r$  and  $\nu$  through the ratio  $r/\nu$ .*

To understand the result stated in Proposition 3, it is instructive to compare the evolution of the net aggregate claim process  $G_{n,1}(\cdot)$  when the client-population-size path is known to be  $F_{n,1}(\cdot) = f(\cdot)$ , to the evolution of the net aggregate claim process  $G_{n,2}(\cdot)$  when the client-population-size path is (say) halved (i.e., it becomes  $F_{n,2}(\cdot) = f(\cdot)/2$ ). Because clients generate claims independently according to a Poisson process  $\nu$  and generate capital at a constant rate  $r$ , we thus have

$$G_{n,1}(t) \stackrel{d}{=} G_{n,2}(2t)$$

for all  $t \geq 0$ . This means that an increase in the number of clients speeds up the evolution of the net aggregate claims (which can be interpreted as time contraction), whereas a decrease in the number of clients slows down the evolution of the net aggregate claims (interpreted as time dilation). This local ‘compressing’ or ‘stretching’ of time evidently has no impact on the probability of eventual bankruptcy. Thus, Proposition 3 reflects the fact that the probability of *eventual* bankruptcy is independent of any contraction/dilation of time. It is noted that the above arguments extend beyond our large deviation context, and therefore imply a more general property of the CL model with fluctuating client population. Indeed, in Proposition 3 the decay rate  $\varrho^*$  can be replaced by ‘the all-time ruin probability’, again relying on the elementary time-contraction/time-dilation argumentation provided above.

Where Proposition 3 concerns the all-time ruin probability, in applications one is, for obvious reasons, typically interested in the time-dependent ruin probability, i.e., the probability of the insurance firm being bankrupt by a given time  $T > 0$ . Importantly, in this case fluctuations in the number of clients *do* play an important role in determining the probability of bankruptcy. As we will show now, however, we can use the ideas that underlie Proposition 3 to identify some structural properties corresponding to this finite-horizon context, too. To this end, let  $(f^{(*,T)}, g^{(*,T)})$  be the most likely path in  $\mathcal{H}_T$ , so that  $\varrho(T) = I_{[0, T]}(f^{(*,T)}, g^{(*,T)})$ . In addition, let  $(\bar{f}, \bar{g})$  satisfy  $I_{[0, \infty)}(\bar{f}, \bar{g}) = 0$ , so that  $(\bar{f}, \bar{g})$  can be interpreted as the fluid limit of  $(\bar{F}_n(\cdot), \bar{G}_n(\cdot))$ .

First consider the case that the horizon  $T$  equals the most likely time  $t^*$  of eventual bankruptcy. In view of the argumentation underlying Proposition 3, one anticipates that the client population evolves (most likely) along its fluid-limit path:

$$(i) \quad f^*(t) := f^{(*,t^*)}(t) = \bar{f}(t) \text{ for all } t \in [0, t^*].$$

Next suppose  $T < t^*$ . In this case the process  $\bar{G}_n(\cdot)$  must enter the rectangular set  $\mathcal{R}$  faster than it would do in the infinite-horizon case. In the most likely path, one thus anticipates that the number of clients is higher than expected in order to speed up the evolution of  $\bar{G}_n(\cdot)$ . This reasoning leads to

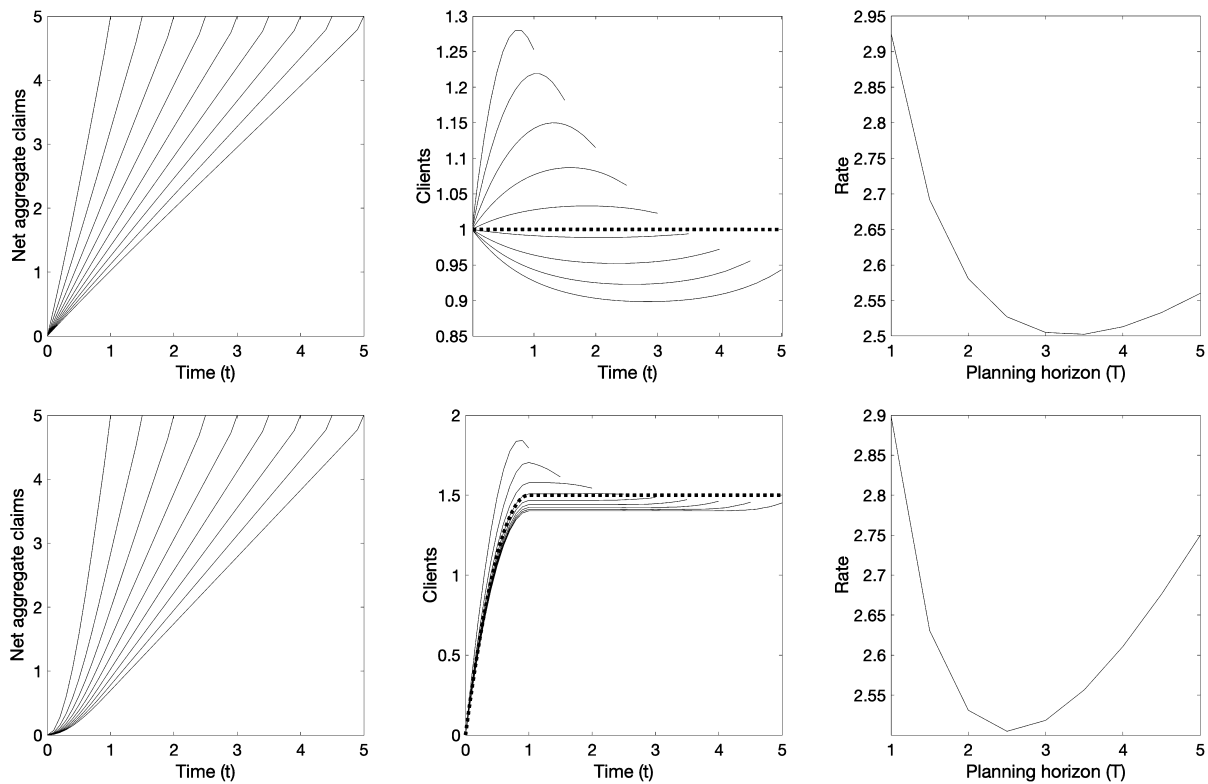
$$(ii) \quad \text{if } T < t^* \text{ then } f^{(*,T)}(t) > f^*(t) = \bar{f}(t) \text{ for all } t \in [0, T].$$

Similarly, if  $T > t^*$ , then  $\bar{G}_n(\cdot)$  must enter  $\mathcal{R}$  more slowly than it would optimally do. In the most likely path, one anticipates the number of clients to be lower than expected in order to slow down the evolution of  $\bar{G}_n(\cdot)$ , i.e.,

$$(iii) \quad \text{if } T > t^* \text{ then } f^{(*,T)}(t) < f^*(t) = \bar{f}(t) \text{ for all } t \in [0, t^*].$$

**Example 1.** We further study the properties (i)–(iii) by means of two numerical experiments that are pictorially illustrated in Fig. 1. In both experiments the net aggregate claim process is characterised by  $\nu = 3, r = 3$ , with the claim sizes being exponentially distributed with mean  $\frac{2}{3}$ . The insurance firm initially has five units of capital (i.e.,  $u = 5$ ), and we consider time horizons  $T \in \{1, 1.5, \dots, 5\}$ . In the top row of Fig. 1 we let  $f_0 = 1, \lambda = 1$ , and the sojourn-time distribution be exponential with mean 1 (with the residual sojourn times of the clients present at time 0 being exponential with mean 1 as well). Note that this means  $\bar{f}(t) = 1$  (dashed curve) for all  $t \geq 0$ ; informally, the population-size process starts in equilibrium. In the bottom row of Fig. 1 we let  $f_0 = 0, \lambda = 3$ , and the sojourn-time distribution be uniform on  $[0, 1]$  (with  $h^\circ(\cdot)$  being the corresponding residual distribution).

Observe that the time horizon associated with the minimal decay rate (i.e., the most likely timescale of ruin in the infinite-horizon case) is  $T = 3.5 \approx t^*$  for the parameter values in the top row, and  $T = 2.5 \approx t^*$  for the parameter values in the bottom row. In the right



**Fig. 1.** Most likely paths to bankruptcy for various values of the time horizon  $T$ . Left panels: most likely net aggregate claim path  $g^{(\star,T)}(\cdot)$ , middle panels: most likely client-population-size path  $f^{(\star,T)}(\cdot)$ , right panels: the corresponding decay rate  $I_{[0,T]}(g^{(\star,T)}, f^{(\star,T)})$ . The top and bottom panels correspond to different parameter values, which are given in the text. The dashed curve in the middle panels is  $\bar{f}(\cdot)$ .

panels of Fig. 1 we see that, in line with Proposition 3, the rates associated with these optimal time horizons are equal (with  $\varrho^\star \approx 2.5$ ). In addition, in the centre column of Fig. 1 we see that, corroborating the properties (i)–(iii) above, the conditioned path of the clients  $f^{(\star,T)}(\cdot)$  is larger than  $\bar{f}(\cdot)$  (depicted by the dashed curve) when  $T < t^\star$ , smaller than  $\bar{f}(\cdot)$  when  $T > t^\star$ , and equal to  $\bar{f}(\cdot)$  when  $T = t^\star$ .

### 3.2.2. What is the primary cause of fluctuations in capital: clients or claims?

We suppose that a *net profit condition* is in place, i.e., we are in the situation that  $r > \bar{m}\nu$ , where  $\bar{m}$  is the expected value of the claim size. This condition is natural as it entails that, on average, each client generates a positive return for the insurance company. Our objective is to understand the most likely cause of unusual values of the net aggregate claim process. Evidently this is connected to ruin, as ruin occurs when the net aggregate claim process is unusually large. However, for ease of exposition we start by considering the case that the net aggregate claim process attains an unusually *small* value  $a$  at time  $T$  (corresponding to an unusually large value of the surplus process).

Suppose  $a < \bar{g}(T)$ , where we recall that  $\bar{g}(\cdot)$  is the fluid limit corresponding to the net aggregate claim process  $G_n(\cdot)$ . One could distinguish between two possible causes for a large surplus to happen. Contribution (1) reflects the event that the number of clients that the insurance company attracts is larger than one would expect; due to the net profit condition this scenario corresponds to a higher surplus. Contribution (2) reflects the event that the client-population size attains its expected value but the amount of money claimed by the clients present is lower than expected. Our objective is to quantify the Contributions (1) and (2). To determine the proportion of the additional capital  $\bar{g}(T) - a$  that can be attributed to additional clients (i.e., Contribution 1) we introduce the performance metric

$$E_1(a, T) := \frac{(r - \bar{m}\nu) \int_0^T [f^{(\star,T)}(t) - \bar{f}(t)] dt}{\bar{g}(T) - a}, \tag{12}$$

with, as before,  $\bar{f}(\cdot)$  denoting the fluid limit of the process  $F_n(\cdot)$ . Observe that the numerator of (12) can be interpreted as the additional clients  $f^{(\star,T)}(t) - \bar{f}(t)$  in the most likely path multiplied by the expected net rate  $r - \bar{m}\nu > 0$  that these clients generate capital, integrated over time. We divide by the total additional capital  $\bar{g}(T) - a$  to obtain a proportion. What remains can be attributed to clients generating fewer claims than expected (i.e., Contribution 2),

$$E_2(a, T) := 1 - E_1(a, T).$$

In this way we have separated the effect due to the fluctuations in the number of clients on one hand, and the effect due to the fluctuations in the amount of money claimed by the clients on the other hand.

Under the net profit condition, when  $a < \bar{g}(T)$  we expect that  $E_1(a, T) > 0$  (and hence  $E_1(a, T) \in [0, 1]$ ). This is due to the time contraction/dilation arguments in Section 3.2.1. In particular, if  $a < \bar{g}(T)$  then there exists  $t' > 0$  such that  $a = \bar{g}(T + t')$ , so that, in order to move toward the optimal time scale, time should contract and hence we expect that  $f^{(\star,T)}(t) > \bar{f}(t)$  for all  $t \in [0, T]$ . The same

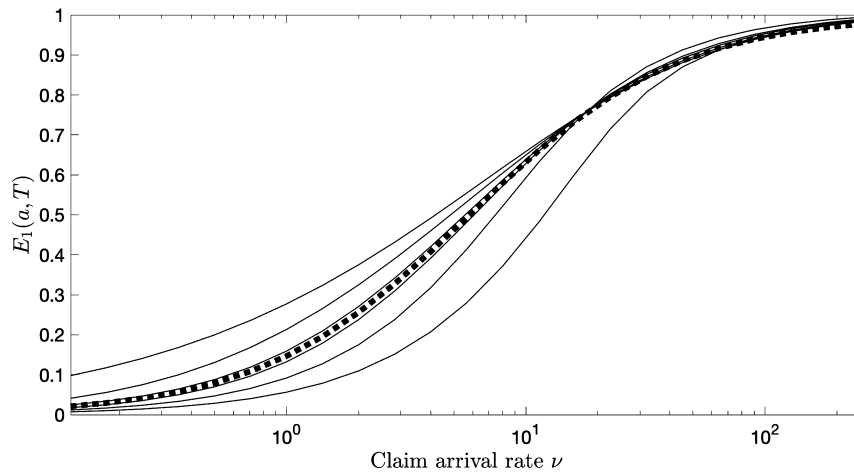


Fig. 2. The proportional effect of fluctuations in the number of clients:  $E_1(a, T)$  as a function of the claim arrival rate  $\nu$ .

reasoning holds when  $a \in [\bar{g}(T), 0]$ , although now with  $t' < 0$  and time dilation; however it breaks down when  $a > 0$ , and in this case we may have  $E_1(a, T) < 0$ .

While  $E_1(a, T)$  and  $E_2(a, T)$  can be computed numerically, in general, it is challenging to express them analytically. However, from Theorem 1 and elementary (but lengthy) calculations we can derive an expression as  $a - \bar{g}(T) \rightarrow 0$ . These calculations, sketched in Appendix B, involve equating the reward (gain in capital) per unit cost (increase in the rate function) for increasing the number of clients that arrive at any time  $t$ , and decreasing the value of claims generated by the clients. In particular, when the sojourn-time distribution is exponential with rate  $\mu$ , we obtain

$$\lim_{a \rightarrow \bar{g}(T)} E_1(a, T) = \frac{\int_0^T (\lambda + \bar{f}(t)\mu) \left[ \frac{r - \nu\beta'(0)}{\mu} (1 - e^{-\mu(T-t)}) \right]^2 dt}{\beta''(0)\nu \int_0^T \bar{f}(t) dt + \int_0^T (\lambda + \bar{f}(t)\mu) \left[ \frac{r - \nu\beta'(0)}{\mu} (1 - e^{-\mu(T-t)}) \right]^2 dt} \tag{13}$$

The individual expressions appearing in the right-hand side of (13) have the following interpretations. In the first place,  $(\lambda + \bar{f}(t)\mu) dt$  is proportional to the variance of the difference in the number of clients at  $t$  and  $t + dt$ , respectively, when the number of clients at time  $t$  is close to the fluid limit  $\bar{f}(t)$ . Secondly,

$$\frac{r - \nu\beta'(0)}{\mu} (1 - e^{-\mu(T-t)})$$

is the expected capital that is earned from a single client that arrives at time  $t$ . Thirdly,

$$\beta''(0)\nu \int_0^T \bar{f}(t) dt$$

is proportional to the variance in the total value of claims when the number of clients follows its fluid limit  $\bar{f}(\cdot)$ . In view of the above,  $\lim_{a \rightarrow \bar{g}(T)} E_1(a, T)$  has the appealing interpretation of a ratio of variances. A similar expression with the same interpretation can be obtained when considering the case with general sojourn times. This expression, being considerably more involved, is left out.

**Example 2.** We illustrate the concepts introduced above by means of a numerical example. We let  $\lambda = 1$ ,  $f_0 = 1$ ,  $T = 1$ , and suppose clients leave the company at rate  $\mu = 1$  (where it is noted that this implies that  $\bar{f}(t) = 1$  for all  $t \in [0, T]$ ). Regarding the claim arrival process, we let  $r = 2$ , and suppose that the claim sizes are exponentially distributed with a mean  $\bar{m}$  such that  $\nu\bar{m} = 1$ . Observe that  $\bar{g}(T) = (\bar{m}\nu - r) \int_0^T \bar{f}(t) dt = -1$ . In Fig. 2 we take values of  $\nu$  ranging from  $2^{-3}$  to  $2^8$  and plot  $E_1(a, T)$  for  $a = 0$  (given by the lowest solid curve),  $-0.5$ ,  $-0.9$ ,  $-1.1$ ,  $-1.5$ ,  $-2$  (given by the highest solid curve), and we plot the limiting value (12) (given by the dashed curve). The figure illustrates that when  $\nu$  is very large (and hence  $\bar{m}$  very small), then  $E_1(a, T)$  is close to 1. This reflects the fact that, under  $\nu\bar{m} = 1$ , as  $\nu \rightarrow \infty$ , each client generates claims in an increasingly deterministic manner, and hence large fluctuations in the capital are more likely to be caused by fluctuations in the number of clients. Evidently, the opposite reasoning applies when  $\nu \downarrow 0$ . The figure also shows that, in the setting considered, for lower values of  $a$ , fluctuations in the number of clients play an increasingly important role.

#### 4. Proofs of the large-deviation principles

##### 4.1. Finite-dimensional LDP

To establish the multi-point LDP stated in Proposition 2 it remains to compute  $M_{\mathbf{t}}^-(\omega, \theta)$  and  $M_{\mathbf{t}}^+(\omega, \theta)$  (as defined in (9)). We start by evaluating  $M_{\mathbf{t}}^-(\omega, \theta)$ .

Recall that  $\tau^\circ$  is a variable corresponding to a typical residual sojourn-time duration of a client who is present at time 0. Let  $\Omega_k := \sum_{j=1}^k \omega_j$ . With  $t_0 \equiv 0$  and, as before,  $\bar{h}^\circ(t) := \int_t^\infty h^\circ(s) ds$ ,



$$\begin{aligned}
 M_{\mathbf{f}}^{-}(\boldsymbol{\omega}, \boldsymbol{\theta}) &= \mathbb{E} \exp \left( \sum_{j=1}^d \omega_j 1\{\tau^\circ > t_j\} + \sum_{j=1}^d \theta_j Z(\tau^\circ \wedge t_j) \right) \\
 &= \sum_{k=1}^d \int_{t_{k-1}}^{t_k} h^\circ(s) \mathbb{E} \exp \left( \sum_{j=1}^d \theta_j Z(s \wedge t_j) \right) \exp(\Omega_{k-1}) \, ds + \tag{14}
 \end{aligned}$$

$$\bar{h}^\circ(t_d) \mathbb{E} \exp \left( \sum_{j=1}^d \theta_j Z(t_j) \right) \exp(\Omega_d). \tag{15}$$

To further evaluate  $M_{\mathbf{f}}^{-}(\boldsymbol{\omega}, \boldsymbol{\theta})$ , let us first focus on Expression (15). By using a telescopic sum representation, and denoting  $\Theta_k := \sum_{j=k}^d \theta_j$  and  $\delta_k := t_k - t_{k-1}$ , we obtain that the mgf featuring in this term equals

$$\mathbb{E} \exp \left( \sum_{j=1}^d \theta_j \sum_{k=1}^j (Z(t_k) - Z(t_{k-1})) \right) = \mathbb{E} \exp \left( \sum_{k=1}^d \Theta_k (Z(t_k) - Z(t_{k-1})) \right) = \prod_{k=1}^d (\varphi(\Theta_k))^{\delta_k}.$$

The other term in  $M_{\mathbf{f}}^{-}(\boldsymbol{\omega}, \boldsymbol{\theta})$ , i.e., Expression (14), can be computed along the same lines. To this end, we use that, evidently, for  $s \in [t_{k-1}, t_k]$  we have that  $t_j \wedge s = t_j$  for  $j = 0, \dots, k-1$ , whereas  $t_j \wedge s = s$  for  $j = k, \dots, d$ . By some standard algebra, we thus obtain that the mgf featuring in the  $k$ -th term in the sum, for  $s \in [t_{k-1}, t_k]$ ,

$$\mathbb{E} \exp \left( \sum_{j=1}^{k-1} \theta_j Z(t_j) + Z(s) \sum_{j=k}^d \theta_j \right) = \left( \prod_{j=1}^{k-1} (\varphi(\Theta_j))^{\delta_j} \right) \cdot (\varphi(\Theta_k))^{s-t_{k-1}}$$

Upon combining the above, we have found the following expression for  $M_{\mathbf{f}}^{-}(\boldsymbol{\omega}, \boldsymbol{\theta})$ . Observe that it is fully in terms of the ‘partial sum series’  $\Theta_k$  and  $\Omega_k$ , corresponding the arguments  $\boldsymbol{\theta}$  and  $\boldsymbol{\omega}$ , respectively.

**Lemma 1.** For  $\boldsymbol{\omega}, \boldsymbol{\theta} \in \mathbb{R}^d$ ,

$$M_{\mathbf{f}}^{-}(\boldsymbol{\omega}, \boldsymbol{\theta}) = \sum_{k=1}^d e^{\Omega_{k-1}} \prod_{j=1}^{k-1} (\varphi(\Theta_j))^{\delta_j} \int_{t_{k-1}}^{t_k} h^\circ(s) \varphi(\Theta_k)^{s-t_{k-1}} \, ds + e^{\Omega_d} \bar{h}^\circ(t_d) \prod_{k=1}^d (\varphi(\Theta_k))^{\delta_k}. \tag{16}$$

We now evaluate  $M_{\mathbf{f}}^{+}(\boldsymbol{\omega}, \boldsymbol{\theta})$ . This is done by distinguishing the contributions due to clients who arrive in each of the intervals  $[t_{\ell-1}, t_\ell)$ , for  $\ell = 1, \dots, d$ , which are independent (as argued before). We thus arrive at the decomposition

$$M_{\mathbf{f}}^{+}(\boldsymbol{\omega}, \boldsymbol{\theta}) = \prod_{\ell=1}^d M_{\mathbf{f},\ell}^{+}(\boldsymbol{\omega}, \boldsymbol{\theta}), \tag{17}$$

where

$$M_{\mathbf{f},\ell}^{+}(\boldsymbol{\omega}, \boldsymbol{\theta}) = \sum_{k=0}^{\infty} e^{-\lambda \delta_\ell} \frac{(\lambda \delta_\ell)^k}{k!} \left( \bar{M}_{\mathbf{f},\ell}^{+}(\boldsymbol{\omega}, \boldsymbol{\theta}) \right)^k = \exp \left( \lambda \delta_\ell \left( \bar{M}_{\mathbf{f},\ell}^{+}(\boldsymbol{\omega}, \boldsymbol{\theta}) - 1 \right) \right), \tag{18}$$

here  $\bar{M}_{\mathbf{f},\ell}^{+}(\boldsymbol{\omega}, \boldsymbol{\theta})$  is the mgf that corresponds to the contribution of a single client arriving at a uniformly distributed epoch in  $[t_{\ell-1}, t_\ell)$ , an interval of length  $\delta_\ell$ . As a consequence, with  $t_{j,\ell} := t_j - t_\ell$ , we have that

$$\bar{M}_{\mathbf{f},\ell}^{+}(\boldsymbol{\omega}, \boldsymbol{\theta}) = \mathbb{E} \exp \left( \sum_{j=\ell}^d \theta_j Z(\tau \wedge (\delta_\ell(1-U) + t_{j,\ell})) + \sum_{j=\ell}^d \omega_j 1\{\tau > \delta_\ell(1-U) + t_{j,\ell}\} \right).$$

By distinguishing between the values of  $\tau$ , the expression in the previous display can be decomposed into the sum of the three terms. The first term corresponds with the scenario that the client has left by time  $t_\ell$ . It can be written as

$$\begin{aligned}
 \bar{M}_{\mathbf{f},\ell,1}^{+}(\boldsymbol{\omega}, \boldsymbol{\theta}) &:= \int_0^{\delta_\ell} \frac{1}{\delta_\ell} \int_0^{\delta_\ell-s} h(r) \mathbb{E} \exp \left( \sum_{j=\ell}^d \theta_j Z(r) \right) \, dr \, ds \\
 &= \int_0^{\delta_\ell} \frac{1}{\delta_\ell} \int_0^{t_{\ell,\ell-1}-s} h(r) (\varphi(\Theta_\ell))^r \, dr \, ds. \tag{19}
 \end{aligned}$$

The second term corresponds to the scenario that the client has left between  $t_k$  and  $t_{k+1}$ , for some index  $k \in \{\ell, \dots, d-1\}$ . We obtain

$$\begin{aligned}
 \bar{M}_{t,\ell,2}^+(\omega, \theta) &:= \int_0^{\delta_\ell} \frac{1}{\delta_\ell} \sum_{k=\ell}^{d-1} \int_{t_{k,\ell-1}-s}^{t_{k+1,\ell-1}-s} h(r) \cdot \\
 &\quad \mathbb{E} \exp \left( \sum_{j=\ell}^k \theta_j Z(t_{j,\ell-1} - s) + \sum_{j=k+1}^d \theta_j Z(r) + \sum_{j=\ell}^k \omega_j \right) dr ds \\
 &= \int_0^{\delta_\ell} \frac{1}{\delta_\ell} \sum_{k=\ell}^{d-1} \int_{t_{k,\ell-1}-s}^{t_{k+1,\ell-1}-s} h(r) e^{\Omega_k - \Omega_{\ell-1}} (\varphi(\Theta_\ell))^{\delta_\ell - s} \cdot \\
 &\quad \left( \prod_{m=\ell+1}^k (\varphi(\Theta_m))^{\delta_m} \right) \cdot ((\varphi(\Theta_{k+1}))^{r - (t_{k,\ell-1} - s)}) dr ds.
 \end{aligned} \tag{20}$$

To verify the expression (20) in the above display, use the distributional equality, with  $Z(s, t) := Z(t) - Z(s)$  for  $s \leq t$ ,

$$\begin{aligned}
 &\sum_{j=\ell}^k \theta_j Z(t_{j,\ell-1} - s) + \sum_{j=k+1}^d \theta_j Z(r) \\
 &\stackrel{d}{=} \Theta_\ell Z(t_{\ell-1} + s, t_\ell) + \sum_{m=\ell+1}^k \Theta_m Z(t_{m-1}, t_m) + \Theta_{k+1} Z(t_k, t_{\ell-1} + s + r),
 \end{aligned}$$

which can be proven by splitting  $Z(t_{j,\ell-1} - s)$  and  $Z(r)$  in the left-hand side into the contributions due to the individual intervals, swapping the order of summation, and using the fact that  $r$  lies in the interval  $[t_{k,\ell-1} - s, t_{k+1,\ell-1} - s)$ , in combination with the fact that all random variables on the right-hand side are independent due to the independent increments property of the Lévy process  $Z(\cdot)$ .

Finally, the third term describes the contribution due to the scenario that the client leaves after  $t_d$ :

$$\begin{aligned}
 \bar{M}_{t,\ell,3}^+(\omega, \theta) &:= \int_0^{\delta_\ell} \frac{1}{\delta_\ell} \int_{t_{d,\ell-1}-s}^\infty h(r) \mathbb{E} \exp \left( \sum_{j=\ell}^d \theta_j Z(t_{j,\ell-1} - s) + \sum_{j=\ell}^d \omega_j \right) dr ds \\
 &= \int_0^{\delta_\ell} \frac{1}{\delta_\ell} \int_{t_{d,\ell-1}-s}^\infty h(r) e^{\Omega_d - \Omega_{\ell-1}} (\varphi(\Theta_\ell))^{\delta_\ell - s} \cdot \left( \prod_{m=\ell+1}^d (\varphi(\Theta_m))^{\delta_m} \right) dr ds.
 \end{aligned} \tag{21}$$

Combining the above, we have found the following expression for  $M_t^+(\omega, \theta)$ . Again it is fully in terms of the ‘partial sum series’  $\Theta_k$  and  $\Omega_k$ .

**Lemma 2.** For  $\omega, \theta \in \mathbb{R}^d$ , we can compute  $M_t^+(\omega, \theta)$  by (17), involving  $\bar{M}_{t,\ell}^+(\omega, \theta)$  via (18). Here  $\bar{M}_{t,\ell}^+(\omega, \theta)$  equals the sum of (19), (20), and (21).

With the moment generating functions given by the expressions above, we thus arrive at the large-deviation principle presented in Proposition 2.

#### 4.2. Sample-path LDP

We now establish the sample-path LDP of Theorem 1. First observe that from the finite-dimensional LDP given in Proposition 2, in combination with the Dawson–Gärtner projective limit theorem (Dembo and Zeitouni, 1998, Thm. 4.6.1), we obtain a sample-path LDP in the pointwise topology (which we denote by  $\mathcal{X}$ ) with rate  $n$  and rate function

$$I_{[0,T]}^{\mathcal{X}}(f, g) := \sup_{d \in \mathbb{N}} \sup_{0 \leq t_1 < \dots < t_d \leq T} I_{(t_1, \dots, t_d)}((f(t_1), \dots, f(t_d)), (g(t_1), \dots, g(t_d))). \tag{22}$$

Recall that  $I_{[0,T]}(f, g)$  is the rate function characterised by (24) and Lemmas 5 and 6 below. To establish Theorem 1 we need to (i) show that

$$I_{[0,T]}^{\mathcal{X}}(f, g) = I_{[0,T]}(f, g),$$

and (ii) strengthen the topology from  $\mathcal{X}$  to the Skorokhod topology by establishing that the sequence of bivariate processes  $\{(\bar{F}_n(\cdot), \bar{G}_n(\cdot))\}_{n \in \mathbb{N}}$  is exponentially tight.

To establish (i) we need to verify that

- (i-a)  $I_{[0,T]}^{\mathcal{X}}(f, g) = \infty$  when  $f$  or  $g$  is not absolutely continuous,
- (i-b)  $I_{[0,T]}^{\mathcal{X}}(f, g) \leq I_{[0,T]}(f, g)$ , and
- (i-c)  $I_{[0,T]}^{\mathcal{X}}(f, g) \geq I_{[0,T]}(f, g)$ .

In Section 4.2.1 we establish properties (i-b) and (i-c) when the rate function is expressed in terms of limiting counterparts of the finite-dimensional mgfs  $M^i(\omega, \theta)$ ,  $i \in \{-, +\}$  that we derived in Section 4.1. We then find explicit expressions for these limiting mgfs. In Section 4.2.2 we prove properties (ii) and (i-a), exploiting the fact that they can be established via similar arguments. This completes the proof of Theorem 1. Finally, in Section 4.2.3 we provide an alternate expression for  $I_{[0,T]}(f, g)$  that may be attractive for computational purposes.

4.2.1. Upper and lower bounds and the limiting mgfs

In our construction we work with a mesh of dimension  $d$  that we make increasingly fine. To this end, we define, for given functions  $\omega(\cdot)$  and  $\theta(\cdot)$ ,

$$d := T/\Delta, \quad t_k := k\Delta, \quad \theta_k := \Delta \theta(k\Delta), \quad \omega_k = \Delta \omega(k\Delta), \quad \text{and} \quad \delta_k := \Delta. \tag{23}$$

In addition, we introduce

$$I_{[0,T]}(f, g) = \sup_{\omega(\cdot), \theta(\cdot)} \left( \int_0^T [\omega(s) f(s) + \theta(s) g(s)] ds - f_0 \log M_{[0,T]}^-(\omega, \theta) - \log M_{[0,T]}^+(\omega, \theta) \right), \tag{24}$$

where the mgfs  $M_{[0,T]}^i(\omega, \theta)$  are given by

$$M_{[0,T]}^i(\omega, \theta) := \lim_{\Delta \downarrow 0} \mathbb{E} \exp \left( \sum_{k=1}^{T/\Delta} \Delta \omega(k\Delta) F_i(k\Delta) + \sum_{k=1}^{T/\Delta} \Delta \theta(k\Delta) G_i(k\Delta) \right), \quad i \in \{-, +\}; \tag{25}$$

the supremum in (24) is taken over all continuous bounded functions on  $[0, T]$ .

**Lemma 3.** *If  $f$  and  $g$  are absolutely continuous, then  $I_{[0,T]}^{\mathcal{X}}(f, g) \geq I_{[0,T]}(f, g)$ .*

**Proof.** We have

$$\begin{aligned} I_{[0,T]}^{\mathcal{X}}(f, g) &\geq \lim_{\Delta \downarrow 0} I_{(t_1, \dots, t_d)}((f(t_1), \dots, f(t_d)), (g(t_1), \dots, g(t_d))) \\ &= \lim_{\Delta \downarrow 0} \sup_{\omega(\cdot), \theta(\cdot)} \left( \sum_{j=1}^d \omega_j f(t_j) + \sum_{j=1}^d \theta_j g(t_j) - f_0 \log M_{\mathbf{t}}^-(\omega, \theta) - \log M_{\mathbf{t}}^+(\omega, \theta) \right) \\ &\geq I_{[0,T]}(f, g), \end{aligned} \tag{26}$$

where the first inequality is due to the definition (22), and where the second inequality follows by contradiction. Indeed, suppose that the inequality does not hold. Then there exists  $\omega^*, \theta^*$  such that

$$\begin{aligned} &\int_0^T [\omega^*(s) f(s) + \theta^*(s) g(s)] ds - f_0 \log M_{[0,T]}^-(\omega^*, \theta^*) - \log M_{[0,T]}^+(\omega^*, \theta^*) \\ &< \lim_{\Delta \downarrow 0} \sup_{\omega(\cdot), \theta(\cdot)} \left( \sum_{j=1}^d \omega_j f(t_j) + \sum_{j=1}^d \theta_j g(t_j) - f_0 \log M_{\mathbf{t}}^-(\omega, \theta) - \log M_{\mathbf{t}}^+(\omega, \theta) \right); \end{aligned}$$

however, if we ignore the supremum on the right-hand side and replace  $(\omega, \theta)$  by  $(\omega^*, \theta^*)$  we obtain equality, i.e., a contradiction.  $\square$

**Lemma 4.** *If  $f$  and  $g$  are absolutely continuous then  $I_{[0,T]}^{\mathcal{X}}(f, g) \leq I_{[0,T]}(f, g)$ .*

**Proof.** Suppose to the contrary that  $I_{[0,T]}^{\mathcal{X}}(f, g) > I_{[0,T]}(f, g)$  for some absolutely continuous  $f$  and  $g$ . In that case there must exist a vector  $\mathbf{t} = (t_1, \dots, t_d)$  such that

$$I_{\mathbf{t}}((f(t_1), \dots, f(t_d)), (g(t_1), \dots, g(t_d))) > I_{[0,T]}(f, g). \tag{27}$$

For  $\ell \in \mathbb{N}$ , let  $\mathbf{s}^\ell = (s_1^{[\ell]}, \dots, s_{k_\ell}^{[\ell]})$  be such that

- for any  $i \in \{1, \dots, d\}$  there exists  $j \in \{1, \dots, k_\ell\}$  with  $t_i = s_j^{[\ell]}$ ,
- $\lim_{\ell \rightarrow \infty} \max_{j \in \{1, \dots, k_\ell\}} |s_j^{[\ell]} - s_{j-1}^{[\ell]}| = 0$ .

By the contraction principle, for any  $\ell \geq 1$ , we have

$$I_{\mathbf{t}}((f(t_1), \dots, f(t_d)), (g(t_1), \dots, g(t_d))) \leq I_{\mathbf{s}^\ell} \left( (f(s_1^{[\ell]}), \dots, f(s_{k_\ell}^{[\ell]})), (g(s_1^{[\ell]}), \dots, g(s_{k_\ell}^{[\ell]})) \right). \tag{28}$$

If we can show that

$$\lim_{\ell \rightarrow \infty} I_{s^{[\ell]}} \left( (f(s_1^{[\ell]}), \dots, f(s_k^{[\ell]})), (g(s_0^{[\ell]}), \dots, g(s_k^{[\ell]})) \right) = I_{[0, T]}(f, g), \tag{29}$$

then, when combined with (28), we have contradicted (27) and hence proved the result. To establish (29) it suffices that we verify that arguments in the proofs of Lemmas 5 and 6 below still apply when

$$\theta_k^{[\ell]} := (s_{k+1}^{[\ell]} - s_k^{[\ell]})\theta(s_k^{[\ell]}), \quad \omega_k^{[\ell]} := (s_{k+1}^{[\ell]} - s_k^{[\ell]})\omega(s_k^{[\ell]})$$

and  $\ell \rightarrow \infty$  (rather than  $\theta_k = \Delta\theta(k\Delta)$  and  $\omega_k = \Delta\omega(k\Delta)$  and  $\Delta \rightarrow 0$ ). As this verification is of a rather mechanical nature, we do not include it here.  $\square$

Our next task is to compute the limiting mgfs  $M_{[0, T]}^-(\omega, \theta)$  and  $M_{[0, T]}^+(\omega, \theta)$ . We start with the (somewhat easier) first mgf, i.e., the one pertaining to  $G_-(\cdot)$  and  $F_-(\cdot)$ . Let

$$\Theta(s) = \int_s^T \theta(r) \, dr, \quad \Omega(s) := \int_0^s \omega(r) \, dr,$$

i.e., the counterparts of the objects  $\Theta_k$  and  $\Omega_k$  that we worked with in the finite-dimensional context, and

$$\Psi_{\omega, \theta}(u) = \Omega(u) + \int_0^u \log \varphi(\Theta(s)) \, ds.$$

**Lemma 5.** For  $\omega \equiv \omega(\cdot)$  and  $\theta \equiv \theta(\cdot)$ ,

$$M_{[0, T]}^-(\omega, \theta) = \int_0^T h^\circ(u) e^{\Psi_{\omega, \theta}(u)} \, du + \bar{h}^\circ(T) e^{\Psi_{\omega, \theta}(T)}.$$

**Proof.** Concerning the second term in the right-hand side of (16), recognising Riemann sums, we readily obtain

$$\begin{aligned} \bar{h}^\circ(T) \lim_{\Delta \downarrow 0} \exp \left( \sum_{k=1}^{T/\Delta} \Delta \omega(k\Delta) \right) \cdot \prod_{k=1}^{T/\Delta} \left( \varphi \left( \sum_{j=k}^{T/\Delta} \Delta \theta(j\Delta) \right) \right)^\Delta \\ = \bar{h}^\circ(T) \lim_{\Delta \downarrow 0} \exp \left( \sum_{k=1}^{T/\Delta} \Delta \left[ \omega(k\Delta) + \log \varphi \left( \sum_{j=k}^{T/\Delta} \Delta \theta(j\Delta) \right) \right] \right) \\ = \bar{h}^\circ(T) \exp \left( \int_0^T \left[ \omega(s) + \log \varphi \left( \int_s^T \theta(r) \, dr \right) \right] \, ds \right) = \bar{h}^\circ(T) e^{\Psi_{\omega, \theta}(T)}. \end{aligned}$$

We continue by focusing on the first term in the right-hand side of (16). We find, again recognising various Riemann sums,

$$\begin{aligned} \lim_{\Delta \downarrow 0} \mathbb{E} \exp \left( \sum_{k=1}^{T/\Delta} \Delta [\omega(k\Delta)F_-(k\Delta) + \theta(k\Delta)G_-(k\Delta)] \right) \\ = \lim_{\Delta \downarrow 0} \sum_{k=1}^{T/\Delta} \int_{(k-1)\Delta}^{k\Delta} h^\circ(s) \varphi \left( \sum_{\ell=1}^{k-1} \Delta \theta(\ell\Delta) \right)^{s-(k-1)\Delta} \, ds \exp \left( \sum_{\ell=1}^{k-1} \Delta \omega(\ell\Delta) \right) \prod_{\ell=1}^{k-1} \left( \varphi \left( \sum_{j=\ell}^{T/\Delta} \Delta \theta(j\Delta) \right) \right)^\Delta \\ = \lim_{\Delta \downarrow 0} \sum_{k=1}^{T/\Delta} \Delta h^\circ(k\Delta) \exp \left( \sum_{\ell=1}^{k-1} \Delta \left[ \omega(\ell\Delta) + \log \varphi \left( \sum_{j=\ell}^{T/\Delta} \Delta \theta(j\Delta) \right) \right] \right) \\ = \int_0^T h^\circ(u) \exp \left( \int_0^u \left[ \omega(s) + \log \varphi \left( \int_s^T \theta(r) \, dr \right) \right] \, ds \right) \, du = \int_0^T h^\circ(u) e^{\Psi_{\omega, \theta}(u)} \, du. \end{aligned}$$

This completes the proof.  $\square$

The next step is to evaluate the mgf  $M_{[0, T]}^+(\omega, \theta)$ . We show that it can be expressed in terms of the objects

$$\Phi_{\omega, \theta}(s) := \int_s^T h(r-s) e^{\Psi_{\omega, \theta}(r) - \Psi_{\omega, \theta}(s)} \, dr \quad \text{and} \quad \bar{\Phi}_{\omega, \theta}(s) := \bar{h}(T-s) e^{\Psi_{\omega, \theta}(T) - \Psi_{\omega, \theta}(s)}.$$

**Lemma 6.** For  $\theta \equiv \theta(\cdot)$  and  $\omega \equiv \omega(\cdot)$ ,

$$M_{[0,T]}^+(\omega, \theta) = \exp \left( \lambda \int_0^T (\Phi_{\omega, \theta}(s) + \bar{\Phi}_{\omega, \theta}(s) - 1) ds \right).$$

**Proof.** We proceed in a similar manner as above (again working with the various quantities that were defined in (23)). In this case, however, the expression for  $M_{\mathbf{t}}^+(\omega, \theta)$ , as was provided in Lemma 2, is considerably more complex. We first analyze the quantities (19), (20), and (21) (under the parametrization given in (23)) when  $\Delta$  is small. The results allow us to compute  $M_{[0,T]}^+(\omega, \theta)$  using (17) and (18). The starting point is that, with  $\mathbf{t}, \theta$  and  $\omega$  as in (23),

$$\bar{M}_{\mathbf{t}, \ell}^+(\omega, \theta) = \exp \left( \lambda \Delta \sum_{\ell=1}^{T/\Delta} (\bar{M}_{\mathbf{t}, \ell}^+(\omega, \theta) - 1) \right) = \exp \left( \lambda \Delta \sum_{\ell=1}^{T/\Delta} \bar{M}_{\mathbf{t}, \ell}^+(\omega, \theta) - \lambda T \right).$$

Then recall that  $\bar{M}_{\mathbf{t}, \ell}^+(\omega, \theta)$  is the sum of (19), (20), and (21).

We start by considering the contribution due to (19). Recall that this corresponds to the case where a client arrives and leaves in the same time interval. When this time interval,  $\Delta$ , is becoming infinitely small, it is expected that this term does not play any role in the arguments to come. Indeed, it is elementary to show that

$$\lambda \Delta \sum_{\ell=1}^{T/\Delta} \bar{M}_{\mathbf{t}, \ell, 1}^+(\omega, \theta) = O(\Delta)$$

as  $\Delta \downarrow 0$ , which justifies leaving it out in the rest of the derivation.

Then focus on the contribution due to (20). This corresponds to the clients who have arrived in the time interval  $[t_{\ell-1}, t_\ell)$  and then leave before  $T$ . We note that the  $s$  in (20) lies between 0 and  $\Delta$ , so that it can be argued that when  $\Delta$  gets small, we can replace it by 0. This concretely means, with  $d = T/\Delta$ ,

$$\begin{aligned} \lim_{\Delta \downarrow 0} \lambda \Delta \sum_{\ell=1}^{T/\Delta} \bar{M}_{\mathbf{t}, \ell, 2}^+(\omega, \theta) &= \lim_{\Delta \downarrow 0} \lambda \Delta \sum_{\ell=1}^{T/\Delta} \sum_{k=\ell}^{d-1} \int_{(k-\ell)\Delta}^{(k-\ell+1)\Delta} h(r) e^{\Omega(Tk/d) - \Omega(T\ell/d)} \cdot \left( \prod_{m=\ell}^k (\varphi(\Theta(Tm/d)))^\Delta \right) dr \\ &= \lim_{\Delta \downarrow 0} \lambda \Delta^2 \sum_{\ell=1}^{T/\Delta} \sum_{k=\ell}^{T/\Delta} h((k-\ell)\Delta) e^{\Omega(k\Delta) - \Omega(\ell\Delta)} \exp \left( \Delta \sum_{m=\ell}^k \log \varphi(\Theta(m\Delta)) \right). \end{aligned}$$

Recognising various Riemann sums, we thus obtain

$$\begin{aligned} \lim_{\Delta \downarrow 0} \exp \left( \lambda \Delta \sum_{\ell=1}^{T/\Delta} \bar{M}_{\mathbf{t}, \ell, 2}^+(\omega, \theta) \right) &= \exp \left( \lambda \int_0^T \int_s^T h(r-s) e^{\Omega(r) - \Omega(s)} \exp \left( \int_s^r \log \varphi(\Theta(u)) du \right) dr ds \right) = \exp \left( \lambda \int_0^T \Phi_{\omega, \theta}(s) ds \right), \end{aligned}$$

with  $\Phi_{\omega, \theta}(s) := \int_s^T h(r-s) e^{\Psi_{\omega, \theta}(r) - \Psi_{\omega, \theta}(s)} dr$ .

We conclude by analyzing the contribution due to (21), describing the impact of the clients who arrive in  $[t_{\ell-1}, t_\ell)$  and remain in the system until time  $T$ . Just as we did for the contribution due to (20), observe that the  $s$  in (21) lies between 0 and  $\Delta$ ; it again requires a standard argument to justify that when  $\Delta \downarrow 0$  we can replace it by 0. More concretely, with  $d = T/\Delta$ ,

$$\lim_{\Delta \downarrow 0} \lambda \Delta \sum_{\ell=1}^{T/\Delta} \bar{M}_{\mathbf{t}, \ell, 3}^+(\omega, \theta) = \lim_{\Delta \downarrow 0} \lambda \Delta \sum_{\ell=1}^{T/\Delta} \int_{(d-\ell)\Delta}^\infty h(r) e^{\Omega(T) - \Omega(T\ell/d)} \cdot \left( \prod_{m=\ell+1}^d (\varphi(\Theta(Tm/d)))^\Delta \right) dr,$$

which can be rewritten as

$$\lim_{\Delta \downarrow 0} \lambda \Delta \sum_{\ell=1}^{T/\Delta} \bar{h}(T - \ell\Delta) e^{\Omega(T) - \Omega(\ell\Delta)} \exp \left( \Delta \sum_{m=\ell+1}^{T/\Delta} \log \varphi(\Theta(m\Delta)) \right).$$

We thus conclude that

$$\begin{aligned} \lim_{\Delta \downarrow 0} \exp \left( \lambda \Delta \sum_{\ell=1}^{T/\Delta} \bar{M}_{t,\ell,3}^+(\omega, \theta) \right) &= \exp \left( \lambda \int_0^T \bar{h}(T-s) e^{\Omega(T)-\Omega(s)} \exp \left( \int_s^T \log \varphi(\Theta(r)) dr \right) ds \right) \\ &= \exp \left( \lambda \int_0^T \bar{\Phi}_{\omega,\theta}(s) ds \right), \end{aligned}$$

with  $\bar{\Phi}_{\omega,\theta}(s) = \bar{h}(T-s) e^{\Psi_{\omega,\theta}(T)-\Psi_{\omega,\theta}(s)}$ .  $\square$

#### 4.2.2. Exponential tightness

To establish exponential tightness we rely on the approach that was developed in Feng and Kurtz (2006). With  $X_n(t) := (F_n(t), G_n(t))$ , we first need a metric  $r$  on  $\mathbb{R}^2$ . To this end, for  $x = (x_1, x_2)$  and  $y = (y_1, y_2)$  define

$$r(x, y) := |x_1 - y_1| + |x_2 - y_2|,$$

and we let  $q(x, y) := r(x, y) \wedge 1$ . Let  $D([0, \infty))$  be the càdlàg space in which the trajectories of  $X_n(\cdot)$  are contained and equip it with the Skorokhod topology. In the sequel  $\{\mathcal{F}_t^n\}_{0 \leq t \leq T}$  is a (naturally chosen) filtration that we detail below. In this case (Feng and Kurtz, 2006, Theorem 4.1) implies the following:

Suppose that

- (A)  $\{X_n(t)\}_{n \in \mathbb{N}}$  is exponentially tight for each  $t \geq 0$  and
- (B) for each  $T > 0$ , there exist random variables  $\gamma_n(\delta, \alpha, T)$ , satisfying

$$\mathbb{E} \left[ e^{n\alpha q(X_n(t+u), X_n(t))} \middle| \mathcal{F}_t^n \right] \leq \mathbb{E} \left[ e^{\gamma_n(\delta, \alpha, T)} \middle| \mathcal{F}_t^n \right] \tag{30}$$

for  $0 \leq t \leq T$  and  $0 \leq u \leq \delta$  such that for each  $\alpha > 0$ ,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[ e^{\gamma_n(\delta, \alpha, T)} \right] = 0. \tag{31}$$

Then  $\{X_n(\cdot)\}_{n \geq 0}$  is exponentially tight in  $D[0, \infty)$ .

Observe that, by Proposition 1 (i.e., the LDP pertaining to a single point in time), it follows that for each given value of  $t \geq 0$  the sequence  $\{X_n(t)\}_{n \in \mathbb{N}}$  is exponentially tight, so that the requirement (A) has been taken care of. Hence, to prove Theorem 1, we have to verify requirement (B), i.e., Condition (30) and Condition (31).

Before we verify Condition (30) and Condition (31), we first discuss the filtration  $\{\mathcal{F}_t^n\}_{0 \leq t \leq T}$ . In view of the proofs to follow, we do so by describing the information that is contained in  $\mathcal{F}_t^n$ . Given  $\mathcal{F}_t^n$  we in the first place know the time that each client arrives to the system up to time  $t$ ; we label these times as  $\tau_1 < \tau_2 < \dots < \tau_{A_n(t)}$ , with  $A_n(t)$  the number of client arrivals until time  $t$ . The  $i$ -th arrival is assigned a sojourn time  $S_i$  and, given  $\mathcal{F}_t^n$ , we know whether  $S_i \leq t - \tau_i$  and if this inequality holds then we know the precise value of  $S_i$ . Given  $\mathcal{F}_t^n$  we in addition know whether or not each individual initially present has left the system and, if so, we know her specific residual sojourn time; the residual time of the  $i$ -th client is represented by  $S_i^\circ$ . We also know the claim sizes and claim arrival times pertaining to all the claims that occurred in the time interval  $(0, t]$ .

◦ *Step I: Construction of  $\gamma_n(\delta, \alpha, T)$  so that Condition (30) is met.* To apply (Feng and Kurtz, 2006, Theorem 4.1), the idea is to identify a random variable  $\gamma_n(\delta, \alpha, T)$  that stochastically dominates

$$\begin{aligned} n\alpha q((F_n(t+u), G_n(t+u)), (F_n(t), G_n(t))) \\ \leq n\alpha (|F_n(t+u) - F_n(t)| \wedge 1) + n\alpha (|G_n(t+u) - G_n(t)| \wedge 1) \end{aligned}$$

for any  $\alpha \geq 0, 0 \leq u \leq \delta, 0 \leq t \leq T$ , and  $\mathcal{F}_t^n$ . To this end, we first find a stochastically dominating random variable for  $|F_n(t+u) - F_n(t)| \wedge 1$ . Recall that  $\{A_n(t) : 0 \leq t \leq T\}$  denotes the arrival process of the clients, i.e., a Poisson process with rate  $\lambda n$ . Observe that the change in the number of clients in the system between times  $t$  and  $t+u$  (with  $u \in [0, \delta]$ ) is dominated by the number of clients who arrived in  $(t, t+\delta)$  plus the number of clients that were served in this time interval. Now let  $\bar{A}_n(\delta) \sim \text{Poi}(n\lambda\delta)$ , and let this random quantity be independent of the client arrival process  $A_n(\cdot)$ . Note that, because  $u \in [0, \delta]$ ,  $\bar{A}_n(\delta)$  stochastically dominates the number of clients who arrive in the interval  $(t, t+u)$  given  $\mathcal{F}_t^n$ . In addition, note that the number of clients who leave the system between times  $t$  and  $t+u$  is dominated by the number that left in  $(t, t+\delta)$ , which is given by

$$V_n(\delta, t) := \sum_{i=1}^{nf_0} \mathbf{1}\{S_i^\circ \in [t, t+\delta]\} + \int_0^t du \sum_{i=1}^{A(t)} \mathbf{1}\{\tau_i = u\} \mathbf{1}\{S_i \in [t-u, t-u+\delta]\}. \tag{32}$$

Note that (32) is a function of  $t$ , whereas the dominating random variable that we must construct  $\gamma_n(\delta, \alpha, T)$  should not depend on  $t$ . We thus dominate (32) by  $\sup_{t \in [0, T]} V_n(\delta, t)$  to obtain a bound that is uniform in  $t \in [0, T]$ . Taking into account both arrivals and departures, we then have, for any  $\mathcal{F}_t^n$ ,

$$\begin{aligned} |F_n(t+u) - F_n(t)| \wedge 1 &\leq \frac{\text{st } \bar{A}_n(\delta) + \sup_{t \in [0, T]} V_n(\delta, t)}{n} \wedge 1 \\ &\leq \beta_n^{(1,1)}(\delta, T) + \beta_n^{(1,2)}(\delta, T) + \beta_n^{(1,3)}(\delta, T) =: \beta_n^{(1)}(\delta, T), \end{aligned}$$

where we define

$$\beta_n^{(1,1)}(\delta, T) := \frac{\bar{A}_n(\delta)}{n}, \quad \beta_n^{(1,2)}(\delta, T) := \frac{1}{n} \sup_{t \in [0, T]} \left\{ \sum_{i=1}^{nf_0} \mathbf{1}\{S_i^\circ \in [t, t + \delta]\} \right\},$$

$$\beta_n^{(1,3)}(\delta, T) := \frac{1}{n} \left( \sup_{t \in [0, T]} \left\{ \int_0^t du \sum_{i=1}^{A(T)} \mathbf{1}\{\tau_i = u\} \mathbf{1}\{S_i \in [t - u, t - u + \delta]\} \right\} \wedge n \right).$$

Now that we have succeeded in identifying a stochastically dominating random variable for the first component  $|F_n(t + u) - F_n(t)| \wedge 1$ , we proceed by identifying a stochastically dominating random variable for the second component  $|G_n(t + u) - G_n(t)| \wedge 1$ . The change in the net aggregate claim process between times  $t$  and  $t + u$  is dominated by the premiums paid by the clients in this time interval plus the claims made by the clients in this time interval. Recalling that  $u \in [0, \delta]$ , the premiums paid by the clients between times  $t$  and  $t + u$  is dominated by  $r\eta_n(\delta, T)$  with  $\eta_n(\delta, T) := \delta(nf_0 + A_n(T))$ , and the sum of the claims made by the clients between times  $t$  and  $t + u$  is dominated by

$$\bar{Y}_n(\delta, T) := \sum_{i=1}^{\bar{A}_n(\delta, T)} Y_i, \tag{33}$$

where  $\bar{A}_n(\delta, T) \sim \mathbb{P}\text{oi}(\nu\eta_n(\delta, T))$  and is conditionally independent of everything else given  $A_n(T)$ , and the  $Y_i$  are iid random variables with mgf  $\beta(\cdot)$ . Hence, for any  $\mathcal{F}_t^n$ ,

$$|G_n(t + u) - G_n(t)| \wedge 1 \stackrel{\text{st}}{\leq} \frac{1}{n} (r\delta(nf_0 + A_n(T)) + \bar{Y}_n(\delta, T)) \wedge 1 =: \beta_n^{(2)}(\delta, T).$$

From the above we conclude that Condition (30) is satisfied if we let

$$\gamma_n(\delta, \alpha, T) := \alpha n \left( \beta_n^{(1)}(\delta, T) + \beta_n^{(2)}(\delta, T) \right).$$

We conclude that we have constructed a random quantity  $\gamma_n(\delta, \alpha, T)$  so that Condition (30) is met.

◦ *Step II: Verifying that  $\gamma_n(\delta, \alpha, T)$  is so that Condition (31) is met.* Now we need to verify Condition (31), i.e., we need to show that, for the constructed  $\gamma_n(\delta, \alpha, T)$ , and for any  $\alpha > 0$ ,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[ e^{\gamma_n(\delta, \alpha, T)} \right] = 0.$$

To this end, first observe that by Hölder's inequality

$$\frac{1}{n} \log \mathbb{E} \left[ e^{\gamma_n(\delta, \alpha, T)} \right] \leq \frac{1}{2n} \log \mathbb{E} \left[ e^{2\alpha n \beta_n^{(1)}(\delta, T)} \right] + \frac{1}{2n} \log \mathbb{E} \left[ e^{2\alpha n \beta_n^{(2)}(\delta, T)} \right].$$

Hence to verify (31) we can separately treat each term in the right-hand side of the previous display. We start by establishing

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[ e^{2\alpha n \beta_n^{(1)}(\delta, T)} \right] = 0. \tag{34}$$

Because  $\beta_n^{(1,1)}(\delta, T)$ ,  $\beta_n^{(1,2)}(\delta, T)$ , and  $\beta_n^{(1,3)}(\delta, T)$  are independent, (34) follows from the following lemma.

**Lemma 7.** For  $i = 1, 2, 3$  it holds that

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[ e^{2\alpha n \beta_n^{(1,i)}(\delta, T)} \right] = 0. \tag{35}$$

*Proof:* We treat  $i = 1, 2$ , and 3 separately. For  $i = 1$  we use the known expression for the Poisson mgf, so as to obtain

$$\frac{1}{n} \log \mathbb{E} \left[ e^{2\alpha n \beta_n^{(1,1)}(\delta, T)} \right] = \frac{1}{n} \log \mathbb{E} \left[ e^{2\alpha \bar{A}_n(\delta)} \right] = \lambda \delta (e^{2\alpha} - 1) \rightarrow 0, \quad \text{as } \delta \downarrow 0. \tag{36}$$

For  $i = 2$  we first observe that

$$\sup_{t \in [0, T]} \sum_{i=1}^{nf_0} \mathbf{1}\{S_i^\circ \in [t, t + \delta]\} \leq \max_{k \in \{0, 1, \dots, T/\delta\}} \sum_{i=1}^{nf_0} \mathbf{1}\{S_i^\circ \in [k\delta, k\delta + 2\delta]\}.$$

Recall that it was assumed that the density  $h^\circ(\cdot)$  of the residual sojourn times  $S_i^\circ$  is uniformly bounded by some finite constant  $C$ . Consequently, for any  $k$ ,

$$\sum_{i=1}^{nf_0} \mathbf{1}\{S_i^\circ \in [k\delta, k\delta + 2\delta]\} \stackrel{\text{st}}{\leq} B \sim \mathbb{B}\text{in}(nf_0, 2C\delta)$$

(where  $\delta$  is sufficiently small to guarantee  $2C\delta < 1$ ). By McDiarmid (1998, Theorem 2.3), which is effectively a Chernoff inequality, we have

$$\mathbb{P}(B \geq nf_0(2C\delta + \varepsilon)) \leq \exp \left\{ -nf_0 \left( (2C\delta + \varepsilon) \log \left( 1 + \frac{\varepsilon}{2C\delta} \right) - \varepsilon \right) \right\}$$

for any  $\varepsilon > 0$ . Upon combining the above bound, we thus conclude that

$$\begin{aligned} \mathbb{E} \left[ e^{2\alpha n \beta_n^{(1,2)}(\delta, T)} \right] &\leq \mathbb{E} \left( \exp \left\{ 2\alpha \max_{k \in \{0, 1, \dots, T/\delta\}} \sum_{i=1}^{nf_0} \mathbf{1}\{S_i^\circ \in [k\delta, k\delta + 2\delta]\} \right\} \right) \\ &\leq e^{2\alpha n f_0(2C\delta + \varepsilon)} + e^{2\alpha n f_0} \frac{T}{\delta} \mathbb{P}(B \geq nf_0(2C\delta + \varepsilon)) \\ &\leq e^{2\alpha n f_0(2C\delta + \varepsilon)} + e^{2\alpha n f_0} \frac{T}{\delta} \exp \left\{ -nf_0 \left( (2C\delta + \varepsilon) \log \left( 1 + \frac{\varepsilon}{2C\delta} \right) - \varepsilon \right) \right\} \\ &\leq 2 \max \left\{ e^{2\alpha n f_0(2C\delta + \varepsilon)}, e^{2\alpha n f_0} \frac{T}{\delta} \exp \left\{ -nf_0 \varepsilon \left( \log \left( 1 + \frac{\varepsilon}{2C\delta} \right) - 1 \right) \right\} \right\}; \end{aligned}$$

in the second inequality we distinguish between the contributions of the events  $\{B < nf_0(2C\delta + \varepsilon)\}$  and  $\{B \geq nf_0(2C\delta + \varepsilon)\}$ , respectively. Consequently,

$$\begin{aligned} \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[ e^{2\alpha n \beta_n^{(1,2)}(\delta, T)} \right] \\ \leq \lim_{\delta \downarrow 0} \max \left\{ 2\alpha f_0(2C\delta + \varepsilon), 2\alpha f_0 - f_0\varepsilon \left( \log \left( 1 + \frac{\varepsilon}{2C\delta} \right) - 1 \right) \right\} = 2\alpha f_0\varepsilon, \end{aligned}$$

where in the final step we observe that for any  $\varepsilon > 0$  the second term in the maximum converges to  $-\infty$  as  $\delta \downarrow 0$ . Since  $\varepsilon$  is an arbitrary constant, we obtain (35) for  $i = 2$  by taking  $\varepsilon \downarrow 0$ .

We conclude with the analysis corresponding to  $i = 3$ . We let  $A_n^*(\cdot)$  be a sequence of Poisson processes on  $\mathbb{R}$  with intensity  $\lambda n$  which are independent of everything else, and for  $b < c$  let  $A_n^*[b, c]$  denote the number of points contained in the interval  $(b, c)$ . Observe that

$$\begin{aligned} \sup_{t \in [0, T]} \left\{ \int_0^t du \sum_{i=1}^{A_n(t)} \mathbf{1}\{\tau_i = u\} \mathbf{1}\{S_i \in [t-u, t-u+\delta]\} \right\} &\stackrel{\text{st}}{\leq} \sup_{t \in [0, T]} A_n^*[t, t+\delta] \\ &\leq \max_{k \in \{0, 1, \dots, T/\delta\}} A_n^*[k\delta, k\delta + 2\delta], \end{aligned} \tag{37}$$

where for simplicity we assume that  $T/\delta$  is an integer. To understand the validity of (37), observe that the departure process of clients when there are initially no clients present is dominated by the departure process of clients when there are initially a stationary number of clients present. Equation (37) then follows by from the known property that the latter is a Poisson process with intensity  $\lambda n$ .

If  $Z$  is a Poisson random variable with mean  $a$  we obtain (via a Chernoff bound; see for instance (Ross, 2011, Example 7.3)) that

$$\mathbb{P}(Z - a \geq x) \leq \exp \left\{ -x \left( \log \left( 1 + \frac{x}{a} \right) - 1 \right) - a \log \left( 1 + \frac{x}{a} \right) \right\}.$$

Thus,

$$\mathbb{P}(A_n^*[0, 2\delta] \geq n(\varepsilon + 2\lambda\delta)) \leq \exp \left\{ -n\varepsilon \left( \log \left( 1 + \frac{\varepsilon}{2\lambda\delta} \right) - 1 \right) - 2n\lambda\delta \log \left( 1 + \frac{\varepsilon}{2\lambda\delta} \right) \right\}.$$

Consequently, for any  $\varepsilon > 0$ ,

$$\begin{aligned} \mathbb{E} \left[ e^{2\alpha n \beta_n^{(1,3)}(\delta, T)} \right] &\leq \mathbb{E} \left( \exp \left\{ 2\alpha \max_{k \in \{0, 1, \dots, T/\delta\}} \{A_n^*[k\delta, k\delta + 2\delta]\} \wedge n \right\} \right) \\ &\leq e^{2\alpha n(\varepsilon + 2\lambda\delta)} + e^{2\alpha n} \frac{T}{\delta} \mathbb{P}(A_n^*[0, 2\delta] \geq n(\varepsilon + 2\lambda\delta)) \\ &\leq e^{2\alpha n(\varepsilon + 2\lambda\delta)} + e^{2\alpha n} \frac{T}{\delta} \exp \left\{ -n\varepsilon \left( \log \left( 1 + \frac{\varepsilon}{2\lambda\delta} \right) - 1 \right) - 2n\lambda\delta \log \left( 1 + \frac{\varepsilon}{2\lambda\delta} \right) \right\} \\ &\leq 2 \max \left\{ e^{2\alpha n(\varepsilon + 2\lambda\delta)}, e^{2\alpha n} \frac{T}{\delta} \exp \left\{ -n\varepsilon \left( \log \left( 1 + \frac{\varepsilon}{2\lambda\delta} \right) - 1 \right) \right\} \right\}. \end{aligned} \tag{38}$$

We then obtain

$$\begin{aligned} \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[ e^{2\alpha n \beta_n^{(1,3)}(\delta, T)} \right] &\leq \lim_{\delta \downarrow 0} \max \left\{ 2\alpha(\varepsilon + 2\lambda\delta), 2\alpha - \varepsilon \left( \log \left( 1 + \frac{\varepsilon}{2\lambda\delta} \right) - 1 \right) \right\} \\ &= 2\alpha\varepsilon. \end{aligned}$$

Since  $\varepsilon$  is an arbitrary constant, we have obtained (35) for  $i = 3$  by taking  $\varepsilon \downarrow 0$ .  $\square$

We continue by analysing the contribution corresponding to  $\beta_n^{(2)}(\delta, T)$ .



**Lemma 8.** *It holds that*

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[ e^{2\alpha n \beta_n^{(2)}(\delta, T)} \right] = 0. \tag{39}$$

*Proof:* We distinguish between two cases:  $A_n(T) \leq nK$  and  $A_n(T) > nK$ , where  $K$  is an arbitrary constant that we will select later to suit our purposes. In addition, we consider two sub-cases when  $A_n(T) \leq nK$ : when  $\bar{Y}_n(\delta, T) \leq nK'$  and when  $\bar{Y}_n(\delta, T) > nK'$ , where  $\bar{Y}_n(\delta, T)$  is as defined in (33), and  $K'$  denoting an arbitrary constant. We have

$$\begin{aligned} \mathbb{E} \left[ e^{2\alpha n \beta_n^{(2)}(\delta, T)} \right] &\leq \mathbb{E} \left[ e^{2\alpha n \beta_n^{(2)}(\delta, T)} \mid A_n(T) \leq nK \right] + e^{2\alpha n} \mathbb{P}(A_n(T) > nK) \\ &\leq \mathbb{E} \left[ e^{2\alpha n \beta_n^{(2)}(\delta, T)} \mid A_n(T) \leq nK, \bar{Y}_n(\delta, T) \leq n\delta K' \right] \\ &\quad + e^{2\alpha n} \mathbb{P}(\bar{Y}_n(\delta, T) > n\delta K' \mid A_n(T) \leq nK) + e^{2\alpha n} \mathbb{P}(A_n(T) > nK). \\ &\leq B_1(\delta, K, K') + B_2(\delta, K, K') + B_3(\delta, K), \end{aligned}$$

where

$$\begin{aligned} B_1(\delta, K, K') &:= e^{2\alpha n \delta (rf_0 + rK + K')}, \\ B_2(\delta, K, K') &:= e^{2\alpha n} \mathbb{P}(\bar{Y}_n(\delta, T) > n\delta K' \mid A_n(T) = nK), \\ B_3(\delta, K) &:= e^{2\alpha n} \mathbb{P}(A_n(T) > nK). \end{aligned}$$

To verify (31) we deal with each terms  $B_1(\delta, K, K')$ ,  $B_2(\delta, K, K')$ , and  $B_3(\delta, K)$  separately. The first term is straightforward: clearly,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log B_1(\delta, K, K_2) = 0$$

for any choice of the parameter values. Also the third term,  $B_3(\delta, K, K')$ , can be dealt with in a direct fashion, relying on Cramér’s theorem for the sum of independent Poisson random variables. In particular, we use the fact that if  $\mathbb{E}(A_n(T)/n) \equiv \lambda T < K$ , then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(A_n(T) \geq nK) = -K \log \frac{K}{\lambda T} + K - \lambda.$$

We now have

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log B_3(\delta, K) = 2\alpha - K \log \frac{K}{\lambda T} + K - \lambda.$$

For any  $\alpha > 0$  we can choose  $K$  large enough to ensure that these terms are negative.

Finally, we analyze the second term  $B_2(\delta, K, K')$ , again applying Cramér’s theorem. First observe that under the condition  $A_n(T) = nK$ ,  $\bar{A}_n(\delta, T)$  has a Poisson distribution with mean  $n\delta v(f_0 + K)$ . Using the thinning property of a Poisson process, we see that  $\bar{Y}_n(\delta, T)$  has the same distribution as

$$\sum_{j=1}^n \sum_{i=1}^{Z_j} Y_{i,j}$$

where the  $\{Y_{i,j}\}_{i,j \in \mathbb{N}}$  are independent random variables with mgf  $\beta(\cdot)$ , and  $\{Z_j\}_{j \in \mathbb{N}}$  are independent Poisson random variables with mean  $\delta v(f_0 + K)$ . For any  $j \in \mathbb{N}$ ,

$$\mathbb{E} \left[ e^{\theta \sum_{i=1}^{Z_j} Y_{i,j}} \right] = \exp \{ \delta v(f_0 + K) (\beta(\theta) - 1) \} := J_\delta(\theta, K). \tag{40}$$

Applying Cramér’s theorem for sums of iid random variables, we obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\bar{Y}_n(\delta, T) > n\delta K' \mid A_n(T) = nK) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n\delta} \sum_{j=1}^n \sum_{i=1}^{Z_j} Y_{i,j} > K' \right) \\ &= -I_\delta(K, K'), \end{aligned}$$

where  $I_\delta(K, K') := \sup_{\theta \in \mathbb{R}} (\theta K' - \log J_\delta(\theta, K))$  is the Legendre transform of (40). Now note that  $I_\delta(K, K') \rightarrow \infty$  as  $\delta \downarrow 0$  for any  $K, K' > 0$ . We thus conclude

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log B_2(\delta, K, K') = \lim_{\delta \downarrow 0} [2\alpha - I_\delta(K, K')] = -\infty.$$

Consequently, for any  $\alpha > 0$  we can choose  $K$  and  $K'$  such that

$$\begin{aligned} \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[ e^{2\alpha n \beta_n^{(2)}(\delta, T)} \right] \\ \leq \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log (3 \max \{ B_1(\delta, K, K'), B_2(\delta, K, K'), B_3(\delta, K) \}) = 0. \end{aligned}$$

We have thus verified the claim.  $\square$

Lemmas 7 and 8 entail that we have verified Condition (31). As we had already verified Condition (30), we have finished the proof of Theorem 1.

Now that we have proven that the bivariate process  $(\bar{F}_n(\cdot), \bar{G}_n(\cdot))$  is exponentially tight, we finish this subsection by showing (i-a).

**Lemma 9.** We have  $I_{[0,T]}^{\mathcal{X}}(f, g) = \infty$  when  $f$  or  $g$  is not absolutely continuous.

**Proof.** Recall that

$$I_t(\mathbf{f}, \mathbf{g}) = \sup_{\omega, \theta} \left( \sum_{j=1}^d \omega_j f_j + \sum_{j=1}^d \theta_j g_j - f_0 \log M_t^-(\omega, \theta) - \log M_t^+(\omega, \theta) \right).$$

It thus suffices to show that if  $f$  or  $g$  are not absolutely continuous then there exist sequences  $\{\mathbf{t}^n\}$ ,  $\{\omega^n\}$ , and  $\{\theta^n\}$  such that  $I_{\mathbf{t}^n}(\mathbf{f}^n, \mathbf{g}^n) \rightarrow \infty$ , where  $f_j^n = f(t_j^n)$  and  $g_j^n = g(t_j^n)$ .

We start with the case that  $f$  is not absolutely continuous. This means that there exists  $\delta > 0$  and  $\{s_1^n < u_1^n \leq \dots \leq s_{k_n}^n < u_{k_n}^n\}$  such that  $\sum_{\ell=1}^{k_n} (u_\ell^n - s_\ell^n) \rightarrow 0$ , while  $\sum_{\ell=1}^{k_n} |f(u_\ell^n) - f(s_\ell^n)| \geq \delta$ . Let  $\mathbf{t}^n = (t_i^n)_{i \in \{1, \dots, 2k_n\}}$  where, for  $\ell \in \{1, \dots, k_n\}$ , we have  $t_{2\ell-1} = s_\ell^n$  and  $t_{2\ell} = u_\ell^n$ . In addition, let  $\omega^n = (\omega_i^n)_{1 \leq i \leq 2k_n}$  where, for  $\ell \in \{1, \dots, k_n\}$ , we have

$$\omega_{2\ell-1} = \alpha(1 - \mathbf{21}\{f(t_{2\ell}) \geq f(t_{2\ell-1})\}) \quad \text{and} \quad \omega_{2\ell} = \alpha(\mathbf{21}\{f(t_{2\ell}) \geq f(t_{2\ell-1})\} - 1),$$

and  $\theta^n = \mathbf{0}$ . Then

$$\sum_{j=1}^{2k_n} \omega_j^n f_j^n = \alpha \sum_{\ell=1}^{k_n} |f(u_\ell^n) - f(s_\ell^n)| \geq \alpha \delta.$$

Since  $\alpha$  is an arbitrary constant the result is proved if we can show that

$$f_0 \log M_{\mathbf{t}^n}^-(\omega^n, \mathbf{0}) + \log M_{\mathbf{t}^n}^+(\omega^n, \mathbf{0}) \rightarrow 0, \quad \forall \alpha > 0.$$

Due to the particular choice of  $\omega^n$ ,  $M_{\mathbf{t}^n}^-(\omega^n, \mathbf{0})$  and  $M_{\mathbf{t}^n}^+(\omega^n, \mathbf{0})$  only capture changes in the client population size during  $\cup_{\ell=1}^{k_n} [u_\ell^n, s_\ell^n]$ . Since this interval is vanishing many of the arguments used to establish exponential tightness carry over (in this case Lemma 7 specifically), and hence we will be brief with our explanations. In particular, we have

$$M_{\mathbf{t}^n}^-(\omega^n, \mathbf{0}) \leq 1 + e^\alpha \mathbb{P} \left( \tau^\circ \in \bigcup_{\ell=1}^{k_n} [u_\ell^n, s_\ell^n] \right) \rightarrow 1 \tag{41}$$

$$M_{\mathbf{t}^n}^+(\omega^n, \mathbf{0}) \leq \exp \left\{ \lambda \sum_{\ell=1}^{k_n} (u_\ell^n - s_\ell^n) (e^{2\alpha} - 1) \right\} \rightarrow 1, \tag{42}$$

where  $\tau^\circ$  is a random variable with density  $h^\circ(\cdot)$ ; the convergence in (41) follows from the existence of a density (recall the setup in Section 2), whereas the inequality in (42) follows from the observation that when clients arrive according to a Poisson process with rate  $\lambda$  then their departures are dominated by a Poisson process with rate  $\lambda$  (recall the explanation after (37)) in combination with Hölder's inequality.

When  $g$  is not absolutely continuous the arguments are similar (i.e., we let  $\theta^n$  play the role of  $\omega^n$  above), but to establish

$$f_0 \log M_{\mathbf{t}^n}^-(\mathbf{0}, \theta^n) + \log M_{\mathbf{t}^n}^+(\mathbf{0}, \theta^n) \rightarrow 0, \quad \forall \alpha > 0$$

we now follow the same line of reasoning that led to Lemma 8.  $\square$

#### 4.2.3. Alternative expression for the action functional

In this subsection we provide an alternative expression for the action functional  $I_{[0,T]}(f, g)$ , which may be attractive for computational purposes. The main idea is that we decompose the action functional based on the observation that  $F_n(\cdot)$  modulates  $G_n(\cdot)$  and as such evolves independently. This informally means that we can write  $I_{[0,T]}(f, g)$  as the action functional describing the cost of  $\bar{F}_n(\cdot)$  being close to  $f(\cdot)$ , increased by the action functional describing the cost of  $\bar{G}_n(\cdot)$  being close to  $g(\cdot)$  conditional on  $\bar{F}_n(\cdot)$  being close to  $f(\cdot)$ . Below we provide expressions for both components featuring in this decomposition. The same type of decomposition has appeared, in for instance (Huang et al., 2016; Liptser, 1996).

We start by evaluating the action functional of  $\bar{F}_n(\cdot)$  for the path  $f(\cdot)$ . Note that, in passing, we established a ‘marginal LDP’ for the client-population size only (i.e., not including the net aggregate claim process). From the joint LDP of the client-population-size process and the net aggregate claim process, we find that the corresponding rate function reads

$$I_{[0,T]}(f) := \sup_{\omega(\cdot)} \left\{ \int_0^T \omega(s) f(s) ds - f_0 \log \left( \int_0^T h^\circ(u) e^{\Omega(u)} du + \bar{h}^\circ(T) e^{\Omega(T)} \right) \right\}$$

$$- \lambda \int_0^T \left( \int_s^T h(r-s) \frac{e^{\Omega(r)}}{e^{\Omega(s)}} dr + \bar{h}(T-s) \frac{e^{\Omega(T)}}{e^{\Omega(s)}} - 1 \right) ds \Big\}.$$

A complication of this optimization problem is that the argument  $\omega(\cdot)$  also appears as its integrated version  $\Omega(\cdot)$ . However, by applying integration by parts,  $\omega(\cdot)$  can be eliminated from this variational problem, so that it is written in terms  $\Omega(\cdot)$  only. Indeed, an equivalent variational problem is

$$I_{[0,T]}(f) = \sup_{\Omega(\cdot)} \left\{ \Omega(T)f(T) - \int_0^T \Omega(s)f'(s) ds - f_0 \log \left( \int_0^T h^\circ(u) e^{\Omega(u)} du + \bar{h}^\circ(T) e^{\Omega(T)} \right) - \lambda \int_0^T \left( \int_s^T h(r-s) \frac{e^{\Omega(r)}}{e^{\Omega(s)}} dr + \bar{h}(T-s) \frac{e^{\Omega(T)}}{e^{\Omega(s)}} - 1 \right) ds \right\},$$

with  $\Omega(0) = 0$ . For ease rewriting  $z(s) := \exp(\Omega(s))$ , this further reduces to

$$I_{[0,T]}(f) = \sup_{z(\cdot)} \left\{ \log z(T)f(T) - \int_0^T \log z(s)f'(s) ds - f_0 \log \left( \int_0^T h^\circ(u) z(u) du + \bar{h}^\circ(T) z(T) \right) - \lambda \int_0^T \left( \int_s^T h(r-s) \frac{z(r)}{z(s)} dr + \bar{h}(T-s) \frac{z(T)}{z(s)} - 1 \right) ds \right\}. \tag{43}$$

Conditional on the path  $f(\cdot)$  describing the evolution of the client-population size, we now focus on the action functional of the reserve process  $G_n(\cdot)$ . Given that  $\bar{F}_n(\cdot)$  is close to  $f(\cdot)$ , a path  $g(\cdot)$  of the reserve process has, between 0 and  $T$ , rate function

$$I_{[0,T]}(g | f) = \int_0^T K_{f(s)}(g'(s)) ds, \quad \text{where } K_x(u) := \sup_{\theta} (\theta u - x\varphi(\theta));$$

this relation can be considered as a version of Mogulskii’s theorem corresponding to the setting of a random walk of which the increments have a deterministically time-varying distribution. (Informally, the rationale behind the expression for  $I_{[0,T]}(g | f)$  is that, by ‘locally applying Cramér’s theorem’, it equals

$$\lim_{\Delta \downarrow 0} \Delta \sum_{i=0}^{T/\Delta} \sup_{\theta} \left( \theta g'(i\Delta) - \log e^{\theta r f(i\Delta)} - \log \left( \sum_{k=0}^{\infty} e^{-\nu f(i\Delta)} \frac{(\nu f(i\Delta))^k}{k!} (\beta(\theta))^k \right) \right);$$

evaluating this Riemann sum yields the expression for  $I_{[0,T]}(g | f)$  that was postulated above.) Then  $I_{[0,T]}(f, g)$  can be computed through the relation

$$I_{[0,T]}(f, g) = I_{[0,T]}(f) + I_{[0,T]}(g | f).$$

**Declaration of competing interest**

There’s no competing interest.

**Data availability**

No data was used for the research described in the article.

**Appendix A. Computational techniques**

In this appendix we describe a numerical method to solve the variational problem described by Equations (10) and (11). Recall that our goal is to find the most likely path in the set

$$\mathcal{H}_t = \{(f, g) : (f(t), g(t)) \in \mathcal{R}\}, \quad \text{where } \mathcal{R} = [0, \infty) \times B,$$

i.e., the path  $(f^*, g^*)$  such that  $I_{[0,T]}(f^*, g^*) = \varrho(t) := \inf_{f, g \in \mathcal{H}_t} I_{[0,T]}(f, g)$ . To this end we write  $f^* \equiv f^*(t)$  and  $g^* \equiv g^*(t)$ , and find  $\omega^*$  and  $\theta^*$  such that

$$\varrho(t) = I_t(f^*, g^*) = \omega^* f^* + \theta^* g^* - N_t(\theta^*, \omega^*),$$

where  $N_t(\omega, \theta) := f_0 \log M_t^- + \log M_t^+(\omega, \theta)$ ,  $I_t(\cdot, \cdot)$  is the rate function of the one-point LDP given in Proposition 1, and  $\omega^*$  and  $\theta^*$  are the optimising values of  $\omega$  and  $\theta$ . We will argue that this computation can be used as the basis for an efficient technique that yields the full most likely path; cf. the results for most likely workload paths in queues fed by many iid sources, as developed in e.g. Wischik, 2001.

Fix  $s \in [0, t]$ . By the contraction principle, applying the bivariate LDP, we have

$$I_t(f^*, g^*) = \inf_{f \geq 0, g \in \mathbb{R}} I_{s,t}((f, f^*), (g, g^*)).$$

We wish to identify the optimising  $f$  and  $g$  in the right-hand-side, which can be interpreted as  $f^*(s)$  and  $g^*(s)$ . The optimising arguments in the definition of  $I_{s,t}((f, f^*), (g, g^*))$  are  $((0, \omega^*), (0, \theta^*))$ :

$$I_{s,t}((f^*(s), f^*), (g^*(s), g^*)) = 0 \cdot f^*(s) + \omega^* f^* + 0 \cdot g^*(s) + \theta^* g^* - N_{s,t}((0, \omega^*), (0, \theta^*)),$$

with  $N_{s,t}(\omega, \theta) := f_0 \log M_{s,t}^-(\omega, \theta) + \log M_{s,t}^+(\omega, \theta)$ . As a consequence,

$$g^*(s) = \left. \frac{\partial}{\partial \theta_1} N_{s,t^*}(\omega, \theta) \right|_{(\omega, \theta) = ((0, \omega^*), (0, \theta^*))},$$

$$f^*(s) = \left. \frac{\partial}{\partial \omega_1} N_{s,t^*}(\omega, \theta) \right|_{(\omega, \theta) = ((0, \omega^*), (0, \theta^*))}.$$

As we can do this for any  $s$ , we have found a way to evaluate the full most likely path.

### Appendix B. The limiting value of $E_1(a, T)$

In this appendix we present the calculations that lead to (13). As indicated in the main text, we consider the case that clients remain at the insurance firm for an exponentially distributed length of time with mean  $1/\mu$ . Note that, due to the exponential sojourn times,  $(F_n(t), G_n(t))_{t \in [0, T]}$  is a Markov process. With

$$I_{t_1, t_2}((f_1, f_2), \mathbb{R}^2) := \inf_{(g_1, g_2) \in \mathbb{R}^2} I_{t_1, t_2}((f_1, f_2), (g_1, g_2)),$$

the rate associated to the client-population-size process is

$$I_{[0, T]}(f^{(\bullet, T)}) = \sum_{i=1}^{T/\Delta} I_{\Delta i, \Delta(i+1)}((f^{(\bullet, T)}(\Delta i), f^{(\bullet, T)}(\Delta(i+1))), \mathbb{R}^2)$$

for any  $0 < \Delta < T$  where we have applied to the contraction principle to obtain equality with the rate of the finite-dimensional LDP, and the Markov property to decompose the rate function of the finite-dimensional LDP into a sum. Let  $\varepsilon = a - \bar{g}(T)$ . For ease of exposition we will tacitly assume that  $a < \bar{g}(T)$ , i.e., there is an unusually large surplus at time  $T$ . Let  $t = \Delta i$  and  $dt = \Delta$ . The additional clients that can be attributed to the interval  $[t, t + dt)$  are

$$dt a(t) := f^{(\bullet, T)}(t + dt) - \bar{f}^{(\bullet, T)}(t + dt),$$

where  $\bar{f}^{(\bullet, T)}(t + dt)$  is the expected client-population size at time  $t + dt$  given the client population is  $f^{(\bullet, T)}(t)$  at time  $t$ . The expected total capital generated by each additional client that arrived in  $[t, t + dt)$  by time  $T$  (in the conditioned process) is approximately

$$\int_0^{T-t} e^{-\mu x} (r - v\bar{m}) dt = \frac{r - v\bar{m}}{\mu} (1 - e^{-\mu(T-t)}),$$

where this approximation holds for small  $dt$  and  $\varepsilon$ , and uses the fact that for  $\varepsilon$  small, clients in the conditioned process generate claims in a similar manner as in the unconditioned process. Consequently, the total capital that can be attributed to the additional clients that arrived in the interval  $[t, t + dt)$  is approximately

$$dt c(t) := dt a(t) \frac{r - v\bar{m}}{\mu} (1 - e^{-\mu(T-t)}). \tag{44}$$

The share of the total rate  $I_{[0, T]}(f^{(\bullet, T)}, g^{(\bullet, T)})$  that can be attributed to these additional clients is

$$I_{t, t+dt}((f^{(\bullet, T)}(t), f^{(\bullet, T)}(t + dt)), \mathbb{R}^2) \approx I_{t, t+dt}((\bar{f}(t), \bar{f}(t + dt) + dt a(t)), \mathbb{R}^2)$$

$$\approx (dt)^2 a(t)^2 \left. \frac{\partial^2}{\partial y^2} I_{(t, t+dt)}((\bar{f}(x), y), \mathbb{R}^2) \right|_{y = \bar{f}(x+dt)} \tag{45}$$

$$\approx \frac{dt a(t)^2}{\lambda + \bar{f}(t)\mu},$$

where the first step requires  $\varepsilon$  to be small, the second step follows from a Taylor expansion and requires  $dt$  to be small, and the final step follows from the fact that the second derivative of the Legendre transform evaluated at its mean is the reciprocal of the variance of the underlying random variable. In view of (44) and (45), the marginal increase in rate per unit capital corresponding to increasing or decreasing the additional clients that arrive in  $[t, t + dt)$  is

$$\begin{aligned} \frac{d}{dc(t)} \left( \frac{a(t)^2}{\lambda + \bar{f}(t)\mu} \right) &= \frac{d}{dc(t)} \left( \frac{c(t)^2}{(\lambda + \bar{f}(t)\mu) \left[ \frac{r-v\bar{m}}{\mu} (1 - e^{-\mu(T-t)}) \right]^2} \right) \\ &= \frac{2c(t)}{(\lambda + \bar{f}(t)\mu) \left[ \frac{r-v\bar{m}}{\mu} (1 - e^{-\mu(T-t)}) \right]^2}. \end{aligned} \tag{46}$$

Now observe that the additional capital that can be attributed to clients generating fewer total claims than expected is

$$b := g^{(\star, T)}(T) - (r - v\bar{m}) \int_0^T f^{(\star, T)}(t) dt,$$

where we recall that  $g^{(\star, T)}(T) = \bar{g}(T) + \varepsilon$ . The rate associated with these reduced total claims is

$$I_{[0, T]}(g^{(\star, T)} | f^{(\star, T)}) = \bar{K}_{f^{(\star, T)}}(g^{(\star, T)}(T)) \tag{47}$$

$$\approx \bar{K}_{\bar{f}}(\bar{g}(T) + b), \tag{48}$$

where

$$\bar{K}_{\bar{f}}(x) = \sup_{\theta} (\theta x - \gamma(\theta)), \quad \text{with } \gamma(\theta) = \exp \left( v \int_0^T f(t) dt (\beta(\theta) - 1) \right). \tag{49}$$

Note that (47) follows by the contraction principle, while (48) uses  $f^{(\star, T)}(\cdot) \approx \bar{f}(\cdot)$  for  $\varepsilon$  small. In addition, (49) follows from that fact that given the client population  $f$  the total number of claims is Poisson with mean  $v \int_0^T f(t) dt$ . For  $\varepsilon$  small we then have

$$\bar{K}_{\bar{f}}(\bar{g}(T) + b) \approx \frac{b^2}{\beta''(0)v \int_0^T \bar{f}(t) dt}, \tag{50}$$

where we again apply a Taylor expansion, and use the fact that the second derivative of a Legendre transform is the reciprocal of the variance of the underlying random variable. In view of (50), the marginal increase in rate per unit capital corresponding to increasing or decreasing  $b$  is

$$\frac{d}{db} \left( \frac{b^2}{\beta''(0)v \int_0^T \bar{f}(t) dt} \right) = \frac{2b}{\beta''(0)v \int_0^T \bar{f}(t) dt} \tag{51}$$

By the optimality of  $(f^{(\star, T)}, g^{(\star, T)})$ , and Equations (46) and (51) we have

$$\frac{2c(t)}{(\lambda + \bar{f}(t)\mu) \left[ \frac{r-v\bar{m}}{\mu} (1 - e^{-\mu(T-t)}) \right]^2} = \frac{2b}{\beta''(0)v \int_0^T \bar{f}(t) dt}, \quad \text{for all } t \in [0, T],$$

so that

$$c(t) = \frac{b(\lambda + \bar{f}(t)\mu) \left[ \frac{r-v\bar{m}}{\mu} (1 - e^{-\mu(T-t)}) \right]^2}{\beta''(0)v \int_0^T \bar{f}(t) dt}, \quad \text{for all } t \in [0, T].$$

Because the total additional capital must be  $\varepsilon$ , we have

$$\varepsilon = b + \int_0^T c(t) dt,$$

so that

$$\lim_{\varepsilon \downarrow 0} E_1(a, T) = \frac{\int_0^T c(t) dt}{b + \int_0^T c(t) dt} = \frac{\int_0^T (\lambda + \bar{f}(t)\mu) \left[ \frac{r-v\bar{m}}{\mu} (1 - e^{-\mu(T-t)}) \right]^2 dt}{\beta''(0)v \int_0^T \bar{f}(t) dt + \int_0^T (\lambda + \bar{f}(t)\mu) \left[ \frac{r-v\bar{m}}{\mu} (1 - e^{-\mu(T-t)}) \right]^2 dt},$$

where we observe that right-hand side equals (13), as desired.

## References

- Albrecher, H., Constantinescu, C., Palmowski, Z., Regensburger, G., Rosenkranz, M., 2013. Exact and asymptotic results for insurance risk models with surplus-dependent premiums. *SIAM Journal on Applied Mathematics* 73, 47–66.
- Arendarczyk, M., Debicki, K., Mandjes, M., 2014. On the tail asymptotics of the area swept under the Brownian storage graph. *Bernoulli* 20, 395–415.
- Asmussen, S., Albrecher, H., 2010. *Ruin Probabilities*. World Scientific, Singapore.
- Azencott, R., 1978. Grande déviations et applications. In: *École d'été de Probabilités de St. Flour VIII*, vol. 774.
- Bahadur, R., Zabell, S., 1979. Large deviations of the sample mean in general vector spaces. *Annals of Probability* 7, 587–621.
- Blanchet, J., Chen, X., Lam, H., 2014. Two-parameter sample path large deviations for infinite-server queues. *Stochastic Systems* 4, 206–249.
- Borovkov, A., 1965. New limit theorems in boundary problems for sums of independent terms. *Selected Translations in Mathematical Statistics and Probability* 5, 315–372.
- Boxma, O., Mandjes, M., 2021. Affine storage and insurance risk models. *Mathematics of Operations Research* 46, 1282–1302.
- Cramér, H., 1930. On the Mathematical Theory of Risk. *Skandia Jubilee*, vol. 4.
- Constantinescu, C., Delsing, G., Mandjes, M., Rojas-Nandayapa, L., 2020. A ruin model with a resampled environment. *Scandinavian Actuarial Journal* 2020, 323–341.
- Debicki, K., Mandjes, M., 2015. *Queues and Lévy Fluctuation Theory*. Springer, New York.
- Dembo, A., Zeitouni, O., 1998. *Large Deviations Techniques and Applications*. Springer, New York NY, United States.
- Deuschel, J.-D., Stroock, D., 1989. *Large Deviations*. Academic Press, Boston MA, United States.
- Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Risk Theory. In: *Modelling Extremal Events*. In: *Applications of Mathematics*, vol. 33. Springer, Berlin.
- Feng, J., Kurtz, T., 2006. *Large Deviations for Stochastic Processes*. American Mathematical Society, Providence RI, USA.
- Gerber, H., 1970. An extension of the renewal equation and its application in the collective theory of risk. *Scandinavian Actuarial Journal*, 205–210.
- Huang, G., Mandjes, M., Spreij, P., 2016. Large deviations for Markov-modulated diffusion processes with rapid switching. *Stochastic Processes and Their Applications* 126, 1785–1818.
- Kyprianou, A., 2006. *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer, New York.
- Kyprianou, A., 2013. *Gerber-Shiu Risk Theory*. Springer, New York.
- Liptser, R., 1996. Large deviations for two scaled diffusions. *Probability Theory and Related Fields* 106, 71–104.
- Lundberg, F., 1903. Approximerad framställning af sannolikhetsfunktionen: Aterförsäkering af kollektivrisker. PhD thesis. Almqvist & Wiksell.
- Lundberg, F., 1926. Försäkringsteknisk riskutjämning: Teori. F. Englund's boktryckeri A.B, Stockholm.
- Mandjes, M., Mannersalo, P., Norros, I., van Uitert, M., 2006. Large deviations of infinite intersections of events in Gaussian processes. *Stochastic Processes and Their Applications* 116, 1269–1293.
- Mandjes, M., Norros, I., Glynn, P., 2009. On convergence to stationarity of fractional Brownian storage. *The Annals of Applied Probability* 18, 1385–1403.
- Mandjes, M., van Uitert, M., 2005. Sample-path large deviations for tandem and priority queues with Gaussian inputs. *The Annals of Applied Probability* 15, 1193–1226.
- Mannersalo, P., Norros, I., 2002. A most probable path approach to queueing systems with general Gaussian input. *Computer Networks* 40, 399–412.
- McDiarmid, C., 1998. Concentration. In: *Probabilistic Methods for Algorithmic Discrete Mathematics*. Springer, Berlin, pp. 195–248.
- Mogulskii, A., 1974. Large deviations in the space  $C(0, 1)$  for sums given on a finite Markov chain. *Siberian Mathematical Journal* 15, 43–53.
- Mogulskii, A., 1976. Large deviations for trajectories of multi dimensional random walks. *Theory of Probability and Its Applications* 21, 300–315.
- Rolski, T., Schmidli, H., Schmidt, V., Teugels, J., 2009. *Stochastic Processes for Insurance and Finance*. Wiley, Chichester.
- Ross, N., 2011. Fundamentals of Stein's method. *Probability Surveys*, vol. 8, pp. 210–293.
- Ross, S., 2002. *Introduction to Probability Models*, Eighth Edition. Academic Press Inc, Cambridge, Massachusetts, United States.
- Shwartz, A., Weiss, A., 1995. *Large Deviations for Performance Analysis*. Chapman & Hall, London, United Kingdom.
- Stabile, G., Torrisi, G., 2010. Risk processes with non-stationary Hawkes claims arrivals. *Methodology and Computing in Applied Probability* 12, 415–429.
- Varadhan, S., 1966. Asymptotic probabilities and differential equations. *Communications on Pure and Applied Mathematics* 19, 261–286.
- Wischik, D., 2001. Sample path large deviations for queues with many inputs. *The Annals of Applied Probability* 11, 379–404.