



**HAL**  
open science

## Lapse risk in life insurance: correlation and contagion effects among policyholders' behaviors

Flavia Barsotti, Xavier Milhaud, Yahia Salhi

► **To cite this version:**

Flavia Barsotti, Xavier Milhaud, Yahia Salhi. Lapse risk in life insurance: correlation and contagion effects among policyholders' behaviors. 2016. hal-01282601v1

**HAL Id: hal-01282601**

**<https://hal.science/hal-01282601v1>**

Preprint submitted on 4 Mar 2016 (v1), last revised 15 Sep 2016 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Lapse risk in life insurance: correlation and contagion effects among policyholders' behaviors

Flavia Barsotti<sup>\*(1)</sup> Xavier Milhaud<sup>†(2)</sup> Yahia Salhi<sup>†(2)</sup>

<sup>(1)</sup>Risk Methodologies, UniCredit Spa, Milan, Italy

<sup>(2)</sup>Univ Lyon, ISFA, LSAF EA2429, F-69007 Lyon, France

March 4, 2016

## Abstract

The present paper extends the existing literature on lapse risk by presenting a flexible way to model the lapse decisions in a life insurance portfolio. Correlation and contagion effects among policyholders are embedded in the modeling and risk margins estimates can be easily obtained under both stable regimes and stress scenarios. The proposed approach integrates the effects of policyholders' behaviors through the definition of a mathematical framework where the lapse intensity follows a dynamic contagion process: an *external component*, the *shot-noise intensity*, is added to the Hawkes-based one ([2], [6]). Contrary to previous works, our shot-noise intensity is *not constant* and the resulting intensity process is *not Markovian*. We study the influence of the interest rates dynamics on policyholders' behaviors and the resulting impact on lapse risk margins. Closed-form expressions for the moments of the lapse intensity are provided, showing how the lapse risk is affected by massive copycat behaviors. A sensitivity analysis studies the lapse risk metrics as function of the model's parameters, while a simulation study compares our results with the ones obtained using standard practices. The numerical outputs highlight a potential strong misestimation of lapses under extreme scenarios with classical stress testing methodologies.

**Keywords:** Surrender, Policyholders' Behavior, Dynamic Contagion Process, Hawkes Process, Lapse Risk, Stress Tests, Correlation, Contagion.

**JEL Classification:** G22, C41, G13.

---

\**Corresponding author.* The views expressed in this paper are those of the author and should not be attributed to UniCredit Group or to the author as representative or employee of UniCredit Group. Contact: F. Barsotti (Flavia.Barsotti@unicredit.eu; Risk Methodologies, Group Financial Risks, Group Risk Management, UniCredit Spa, Piazza Gae Aulenti, Tower A, Floor 20th, 20154, Milano, Italy).

<sup>†</sup>The work is supported by both the BNP Paribas Cardif Chair "Data Analytics & Models for Insurance" and the ANR Research Project ANR-13-BS01-0011. The views expressed herein are the authors' owns and do not reflect those endorsed by BNP Paribas. Contacts: X. Milhaud (xavier.milhaud@univ-lyon1.fr), Y. Salhi (yahia.salhi@univ-lyon1.fr).

# 1 Introduction

Since the 1980s<sup>1</sup>, the lapse risk has become one of the three main risks faced by life insurance companies (with market risk and credit risk<sup>1</sup>). This is the reason why an accurate modeling of the lapse rate dynamics is crucial for both risk managers and regulators. Regulators need to understand the lapse dynamics in order to properly define adequate capital requirements; risk managers need to monitor the lapse dynamics in order to properly capture the real risk embedded in their life insurance contracts and underlying their exposure to potential massive surrenders. A proper measurement of lapse risk can in fact prevent the company to incur huge losses and/or liquidity needs. As observed by [14], lapses can affect the risk management activity in a life-insurance company from three main viewpoints: 1) the insurer may suffer huge losses from unexpected lapsed contracts due to upfront investments for the business activity; 2) the insurer may face the loss of potential future profitability opportunities due to lapsed contracts; 3) the insurer may experience adverse selection w.r.t. mortality/morbidity of policyholders. The key aspects are thus the *pricing* and *hedging* of the *surrender option* embedded in most of life insurance contracts. Knowing that this option can be exercised at any time, it is fundamental to consider the dynamic aspect of policyholders' behaviors in the lapse risk modeling. Not only the amount of surrenders is impacting the company's exposure to lapse risk but also a *timing risk* component, due to the American-style nature of these implicit options.

*How to model lapse dynamics* has been addressed in various papers (e.g. [20], [13], [12], [18]); and statistical methodologies include classification trees, generalized linear models, and survival analysis. However, these techniques globally fail at giving accurate individual lapse predictions since policyholders' behaviors are very *heterogeneous*. Indeed, it is well-known that the lapse dynamics depends on many risk factors such as: tax relief, financial markets, contract features, firm's reputation, competition, and regulation. In P&C insurance, for instance, the best explanatory variable is known to be the price elasticity due to yearly renewable contracts ([8]). In life business, the question seems instead more complex, since the contracts remain in force for years. This is crucial to understand why life risk managers are used to model lapses as the result of two components: *structural lapses* (the baseline risk, depending on policyholder's characteristics) and *temporary lapses* (related to short-term disturbances of the environment, financial incentives). In the same spirit, [19] has shown through empirical studies that the most relevant causes of lapses come from two main economic variables: *liquidity needs* for personal projects (e.g. purchase a new car), and *agents' rationality*. Despite the fact that policyholders' personal plans are very difficult to anticipate, in practice it has been observed that they remain quite stable and independent events over time in a large portfolio. This is the reason why risk managers mainly focus on the implications of agents rationality. If agents are rational, a changing environment can make policyholders' decisions become *correlated* and, on top of this, *contagion* effects can be observed in the market (following some experts' recommendations, largely broadcasted rumors, ...). Correlation and contagion effects among policyholders' behaviors could have dramatic consequences on the life insurance industry's exposure to lapse risk since the main assumption underlying every classical statistical model used in practice is the *independence among individuals*. When the markets show a breakdown of this assumption, standard practices for pricing and reserving are not suited to properly capture the real risk faced by insurance companies. And when it arrives, it should already be too late to adjust the model without facing relevant losses. Defining stress scenarios under i) sudden (interest) rates

---

<sup>1</sup>See EIOPA Report on the fifth Quantitative Impact Study (QIS5) for Solvency II, 2011.

movements and ii) the breakdown of behavioral assumptions is one of the main requirements also for banks to be compliant to the most recent EBA regulation for the management of interest rate risk<sup>2</sup>. Copycat behaviors have been observed in life insurance, e.g. in Japan in the 2000s, and represent the major threat for risk managers under whatever potential future scenario: no lapses or massive lapses could be both potentially dangerous, depending on the product guarantees and the economic context. The riskiness increases for the company since the policyholders *simultaneously* take the same decision before the contractual maturity. *Massive surrenders* can thus be observed at any point in time (*timing risk*) before the natural expiring date of the contracts.

To monitor the lapse risk, regulators have historically defined simple risk management rules to calculate the *Solvency Capital Requirement* (SCR): in general, computing the SCR requires to assess a baseline risk to which arbitrary shocks are applied (see the technical specifications of the 5<sup>th</sup> QIS in the European directive Solvency II<sup>3</sup>). For instance, in life insurance and to compute the SCR related to the lapse risk, the upper-shock consists in adding temporary lapses to the structural part, where the temporary lapse rate can represent up to 30% of the whole portfolio exposure. On the other side, when building an internal model, most practitioners currently use a standard deterministic model based on a *S-shaped* curve<sup>4</sup> that links the lapse rate to the policyholder's satisfaction. In [16], the authors suggest a stochastic extension to this framework by modeling insured' decisions with a common shocks model, resulting in a bimodal lapse distribution to account for this 0-1 decision in adverse market scenarios.

Although trying to integrate (artificially or not) the correlation effect that can potentially arise among policyholders' behaviors, these approaches have some limits in that they are static models. Shocks are determined *a priori*, and the potential contagion among policyholders' behaviors in extreme situations is not considered. The present paper extends the existing literature on lapses by presenting a flexible way to model the lapse decisions in a life insurance portfolio: correlation and contagion among policyholders are embedded in the modeling and risk margins estimates can be easily obtained also under stress scenarios. The main novelty of the proposed approach is to integrate *correlation and contagion effects* among policyholders' behaviors into the dynamics of the lapse intensity process. This technique represents a suitable outlet to model the lapse rate dynamics under both unstressed and stressed scenarios, by keeping a full analytical tractability of the relevant risk measures. The proposed methodology leverages on a specific extension of Hawkes contagion processes, firstly introduced in [11]. Hawkes processes are a powerful tool applied in finance and insurance ([10], [4, 5]), based on a piecewise deterministic intensity (see [7]). Hereafter, an *external component*, the *shot-noise intensity*, is added to the Hawkes-based intensity ([2], [6]). This *external component* aims at capturing the correlation among policyholders' behaviors, when reacting to changes in market interest rates under stress scenarios. Our shot-noise intensity is not assumed to be constant and is instead derived from an inverse Gaussian distribution: the resulting lapse intensity process is then not Markovian. The main contribution of the paper thus lies in

---

<sup>2</sup>See the "Final Report" entitled *Guidelines on the management of interest rate risk arising from non-trading activities*, issued by the European Banking Authority (EBA) in May 2015.

<sup>3</sup>More precisely concerning the lapse risk, see [http://archive.eiopa.europa.eu/fileadmin/tx\\_dam/files/consultations/QIS/QIS5/QIS5-technical\\_specifications\\_20100706.pdf](http://archive.eiopa.europa.eu/fileadmin/tx_dam/files/consultations/QIS/QIS5/QIS5-technical_specifications_20100706.pdf), pp.155-159

<sup>4</sup>See the ONC document by the french regulator (ACPR): [https://acpr.banque-france.fr/fileadmin/user\\_upload/acp/International/Les\\_grands\\_enjeux/Exercice-preparation-solvabilite-II/20130527-ONC-2013.pdf](https://acpr.banque-france.fr/fileadmin/user_upload/acp/International/Les_grands_enjeux/Exercice-preparation-solvabilite-II/20130527-ONC-2013.pdf). Note that ACPR does not encourage everyone to use it (to avoid systemic risk).

the definition of a flexible theoretical setting integrating correlation and contagion effects among policyholders behaviors. A sensitivity analysis shows the main lapse risk metrics as function of the model’s parameters, while a simulation study compares our results with those obtained using standard practices. The numerical outputs highlight a potential strong misestimation of lapses under extreme scenarios using classical stress testing methodologies.

The paper is organized as follows. Section 2 discusses the financial setting and describes the proposed mathematical framework by defining the intensity process. Section 3 provides closed-form expressions for the moments of the lapse intensity. Section 4 shows the sensitivity analysis performed in order to determine the qualitative impact of the model parameters on the mean risk level. Finally, numerical results on risk margins and capital requirements are given in Section 5 through simplified real-life examples by comparing our methodology with standard practices (Solvency II, S-shaped curve). Section 6 provides some concluding remarks and Appendix A contains the proofs of our main theoretical results.

## 2 The mathematical model for the lapse dynamics

As already mentioned, lapses are provoked either by structural or temporary risk factors. Structural risk factors indicate the drivers arising from specific policyholders’ needs or taxation, while temporary risk factors are more often related to macroeconomic conditions. The proposed methodological framework aims at modeling the policyholders’ propensity to lapse by means of a *dynamic contagion process* ([6]), in order to build an integrated setting allowing to capture both structural and temporary drivers.

For the sake of simplicity, assume that the only temporary risk factor is the market interest rate level. This makes sense from a business viewpoint, because the interest rate level is widely recognized as a key underlying variable of the firm’s exposure to the lapse risk. In the current financial context of very low interest rates, a sudden increase of interest rates would represent a real risk faced by insurance companies<sup>5</sup> since new (more interesting) opportunities would probably appear in the market. Indeed, observed temporary lapses can be triggered by policyholders’ behaviors in reaction to interest rates movements, and a structured stress testing framework is required so as to monitor the impact of policyholders’ behaviors on specific company’s risk metrics.

More concretely, let us consider a life insurance portfolio of individuals holding saving contracts with guaranteed return  $R^g$ , embedding a surrender option that can be exercised at any time. All along the contract lifetime, policyholders usually compare their *contract credited rate* (further denoted with  $(R_t^c)_{t \geq 0}$ ) to some benchmark rate observed in the market (e.g. a long-term interest rate, further denoted by  $(r_t)_{t \geq 0}$ ). The credited rate encompasses a potential profit benefit depending on the company’s profitability, and can sometimes be higher than the guaranteed rate. The economic intuition underlying the model is that lapse decisions can be driven by the comparison between the *contract credited rate* and the interest rate level<sup>6</sup>. How the risk free interest rate trajectory  $r_t$  can impact on the lapse risk? A rational policyholder is obviously more inclined to lapse as the spread

---

<sup>5</sup>See [https://www.banque-france.fr/fileadmin/user\\_upload/banque\\_de\\_france/publications/Revue\\_de\\_la\\_stabilite\\_financiere/etud4\\_0605.pdf](https://www.banque-france.fr/fileadmin/user_upload/banque_de_france/publications/Revue_de_la_stabilite_financiere/etud4_0605.pdf)

<sup>6</sup>In the following sections, the dynamics of  $(r_t)_{t \geq 0}$  is given as the solution of a stochastic differential equation.

between the interest rate and the credited rate increases. Let us indicate this spread at time  $t = 0$  as the first relative gap  $RG_t^0$ , defined as:

$$RG_t^0 := \frac{r_t - R_0^c}{R_0^c}, \quad (1)$$

where  $r_t$  is the risk-free interest rate at time  $t$ , and  $R_0^c := R^g$  coincides with the guaranteed rate  $R^g$  (close to 0 for most of life insurers in 2015). A large relative gap  $RG_t^0$  indicates that alternative investment opportunities become more attractive than the current insurance position held by the insured. At the contract inception,  $r_t$  at least equals  $R^g$  since the life insurance contract offers more guarantees than a sole risk-free return. We thus have  $RG_0^0 \geq 0$ , with  $RG_0^0 \approx 0$  for market clearing competitive reasons. The same relative gap can be defined for each point in time, and the company can adjust the contract credited rate to prevent massive copycat behaviors. The economic intuition is based on policyholders' reactions towards the profitability of alternative investment opportunities: if the relative gap at time  $t$  becomes greater than a certain level, the insurer can reasonably expect to experience some *temporary surrenders*. Assume that subsequent adjustments of the contractual credited rate can be made by the insurer each time the relative gap becomes greater than an exogenous threshold  $B > 0$  (generalization of this setting to the case of a decreasing interest rate regime would be quite straightforward).

Let us assume  $\hat{T}_1$  being the first time the relative gap hits the exogenous threshold  $B$ , and say that the insurer instantaneously updates the contract credited rate by rising it up to the market interest rate level, i.e.  $R_{\hat{T}_1}^c = r_{\hat{T}_1}$ . As a consequence, the new relative gap  $RG_t^1$  is given by

$$RG_t^1 = \frac{r_t - R_{\hat{T}_1}^c}{R_{\hat{T}_1}^c} = \frac{r_t - r_{\hat{T}_1}}{r_{\hat{T}_1}}, \quad \hat{T}_1 \leq t < \infty.$$

If the insurer follows this mechanism to set the contract credited rate, subsequent adjustments will be operated as soon as  $RG_t^1 = B$ ,  $RG_t^2 = B$ , the generic  $RG_t^k = B$ , and so on. These events thus characterize the sequence  $(\hat{T}_j)_{j=0,1,\dots}$  such that

$$\hat{T}_{j+1} = \inf\{t > \hat{T}_j, RG_t^j = B\}, \quad (2)$$

with the *generic relative gap* defined as

$$RG_t^j = \frac{r_t - R_{\hat{T}_j}^c}{R_{\hat{T}_j}^c} = \frac{r_t - r_{\hat{T}_j}}{r_{\hat{T}_j}}, \quad \hat{T}_j \leq t < \infty. \quad (3)$$

and  $\hat{T}_0 = 0$ . Figure 1 illustrates the updating mechanism of the credited rate for a simulated path of the interest rate  $r_t$ , whose dynamics follows a geometric Brownian motion (GBM) as:

$$\frac{dr_t}{r_t} := \mu dt + \sigma dW_t, \quad r_0 > 0, \quad (4)$$

with  $\mu, \sigma \in \mathfrak{R}^+$  and  $(W_t)_{t \geq 0}$  a standard Brownian motion.

The adjustments of the credited rate are not instantaneous in practice, thus we deem reasonable and more realistic to consider that the propensity of policyholders to lapse can jump. In particular,

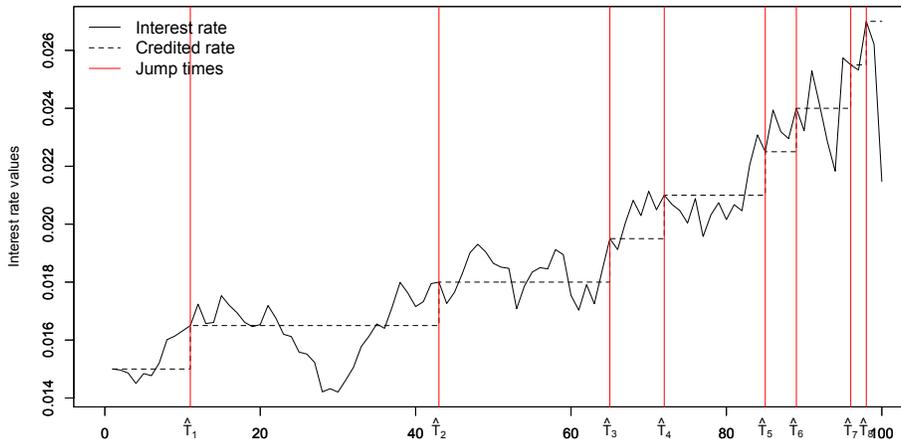


Figure 1: **Market interest rate dynamics and contract credited rate adjustments.** Simulated path of a GBM with  $r_0 = 1.5\%$ ,  $\mu = 5\%$ ,  $\sigma = 30\%$ . The dates  $\hat{T}_j$  at which the contract credited rate  $R_t^c$  is adjusted are reported. Here, the exogenous threshold  $B$  equals 10%.

we suppose it to jump at dates  $\hat{T}_j$ , and then to vanish over time thanks to subsequent efforts made by the insurer. Let  $(N_t)_{t \geq 0}$  be the counting process describing the number of lapses over the whole portfolio at time  $t$ , where lapses occur at dates  $(T_i)_{i=1,2,\dots}$ . Define the intensity  $\lambda_t$  associated to  $N_t$ :

$$\lambda_t = \lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t} + \sum_{i \geq 1} X_i e^{-\beta(t-T_i)} \mathbf{1}_{\{T_i \leq t\}} + \sum_{j \geq 1} Y_j e^{-\beta(t-\hat{T}_j)} \mathbf{1}_{\{\hat{T}_j \leq t\}}, \quad (5)$$

where  $\lambda_0, \lambda_\infty, \beta$  are real positive constants;  $X_i, Y_j$  are random variables for the magnitudes of jumps.

The model in Equation (5) corresponds to a dynamic contagion process (see [6]), i.e. an extension of the classical Hawkes intensity. By using the same notation, the classical Hawkes intensity is given as:

$$\lambda_t = \lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t} + \sum_{i \geq 1} X_i e^{-\beta(t-T_i)} \mathbf{1}_{\{T_i \leq t\}}. \quad (6)$$

Differently from a deterministic Cox-type constant intensity which has been widely used in the actuarial literature (e.g. [3] for lapses), Hawkes-like intensity is stochastic and increases whenever the point process  $N_t$  jumps. These *self-excited jumps* lead to natural contagion and allow to capture massive lapses generated by the breakdown of classical behavioral assumptions. A widespread panic situation in the market could make policyholders much more inclined to lapse due to correlation and contagion effects, and thus making lapse risk much more difficult to monitor ([17]) for insurance companies.

When looking at Equation (6), the contract lapsation at time  $T_i$  has an impact of magnitude  $X_i$  on the intensity process  $\lambda_t$ . This impact exponentially vanishes over time at the (constant) rate  $\beta$  until the mean-reversion level  $\lambda_\infty$  is reached. However, the temporary risk factor that can originally cause the contagion is not explicitly incorporated, while this is the case in Equation (5) thanks to the external jumps at dates  $(\hat{T}_j)$ . In our setting, parameters  $\lambda_0$  and  $\lambda_\infty$  respectively account for initial and long-term hazard rates related to *structural lapses*. Section 5 provides some

explanations concerning the calibration of these parameters based on empirical data.

Without loss of generality, we assume that the magnitudes of *self-excited* and *external* jumps are exponentially-distributed with, respectively, parameters  $\gamma$  and  $\delta$ . The lapse intensity process  $\lambda_t$  therefore depends on both (i) the point process  $N_t$  (i.e. introduced as technical tool to model contagion inside the portfolio), and (ii) another *external* point process  $(\hat{N}_t)_{t \geq 0}$  related to the jumps dates  $\hat{T}_j$ :  $\hat{N}_t = \sum_{j \geq 1} \mathbf{1}_{\{\hat{T}_j \leq t\}}$ . The mixture of these two risk sources provides a suitable framework to integrate and capture in the modeling both contagion and correlation effects observed among policyholders' behaviors. Figure 2 shows a typical trajectory of our intensity process  $\lambda_t$  in Equation (5), corresponding to the interest rate trajectory depicted in Figure 1. Notice that  $N_t$  does not jump at  $\hat{T}_j$ : the only quantity impacted by the external jumps is  $\lambda_t$  itself.

### 3 Theoretical results: moments of the lapse intensity

This section is devoted to the derivation of theoretical results concerning the moments of the lapse intensity  $\lambda_t$  defined in Equation (5). The expected number of lapses within a given time horizon will also be given in closed form. First, let us briefly recall some useful properties of the counting process  $(\hat{N}_t)_{t \geq 0}$ , and the external events dates  $\hat{T}_j$  associated to jumps in the intensity process.

#### 3.1 External jumps

The event dates  $(\hat{T}_j)_{j=0,1,\dots}$  are hitting times of the process describing the evolution of the relative gaps  $RG_t^j$  defined in Equation (3). Our methodology is developed under the assumption of a specific stochastic dynamics for the interest rate. Here, a GBM with drift  $\mu$  and volatility  $\sigma$  is considered (see Equation (4)), and the solution is given by

$$r_t = r_0 \cdot e^{(\mu - \sigma^2/2)t + \sigma W_t}. \quad (7)$$

The generic relative gap  $RG_t^j$  thus satisfies  $\log(RG_t^j + 1) = \mu(t - \hat{T}_j) + \sigma(W_t - W_{\hat{T}_j})$ , and the events  $\hat{T}_j$  can be characterized as follows:

$$\hat{T}_j = \hat{T}_{j-1} + \inf\{t > 0, (\mu - \sigma^2/2)t + \sigma W_t = \log(1 + B)\}, \quad (8)$$

by exploiting the independence of the Brownian increments. This property guarantees that the inter-arrival times  $\Delta \hat{T}_j = \hat{T}_j - \hat{T}_{j-1}$  are independent, with a well known distribution ([24]). Indeed,  $(\Delta \hat{T}_j)_{j=1,2,\dots}$  are inverse Gaussian random variables with mean  $\theta_1 = 2 \log(1 + B) / (2\mu - \sigma^2)$  and shape  $\theta_2 = (\log(1 + B))^2 / \sigma^2$ . Hence, for  $t > 0$ , their density and cumulative distribution function (CDF) are respectively given in closed form by

$$f(t) = \left( \frac{\theta_2}{2\pi t^3} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\theta_2(t - \theta_1)^2}{2\theta_1^2 t} \right\}, \quad (9)$$

$$F(t) = \Phi \left( \sqrt{\frac{\theta_2}{t}} \left( \frac{t}{\theta_1} - 1 \right) \right) + \exp \left\{ \frac{2\theta_2}{\theta_1} \right\} \Phi \left( -\sqrt{\frac{\theta_2}{t}} \left( \frac{t}{\theta_1} + 1 \right) \right), \quad (10)$$

with  $\Phi(\cdot)$  being the standard normal CDF (see [23, p. 43]).

Notice also that  $\hat{T}_j = \sum_{k=1}^j \Delta \hat{T}_k$ , and introduce the renewal function  $h(t) = \mathbb{E}[\hat{N}_t]$ . Knowing that the  $\Delta \hat{T}_j$ 's are independent and identically distributed (i.i.d.), we have  $\mathbb{P}(\hat{T}_j \leq t) = F^{j*}(t)$  where  $F^{j*}$  is the  $j$ -fold convolution of  $F$ . Using  $\mathbb{E}[\hat{N}_t] = \sum_{j=0}^{\infty} \mathbb{P}(\hat{N}_t \geq j)$ , we have

$$h(t) = \mathbb{E}[\hat{N}_t] = \sum_{j=0}^{\infty} F^{j*}(t). \quad (11)$$

As shown in [23], the CDF  $F^{j*}$  is still an inverse Gaussian, with mean  $j\theta_1$  and shape  $j\theta_2$ .

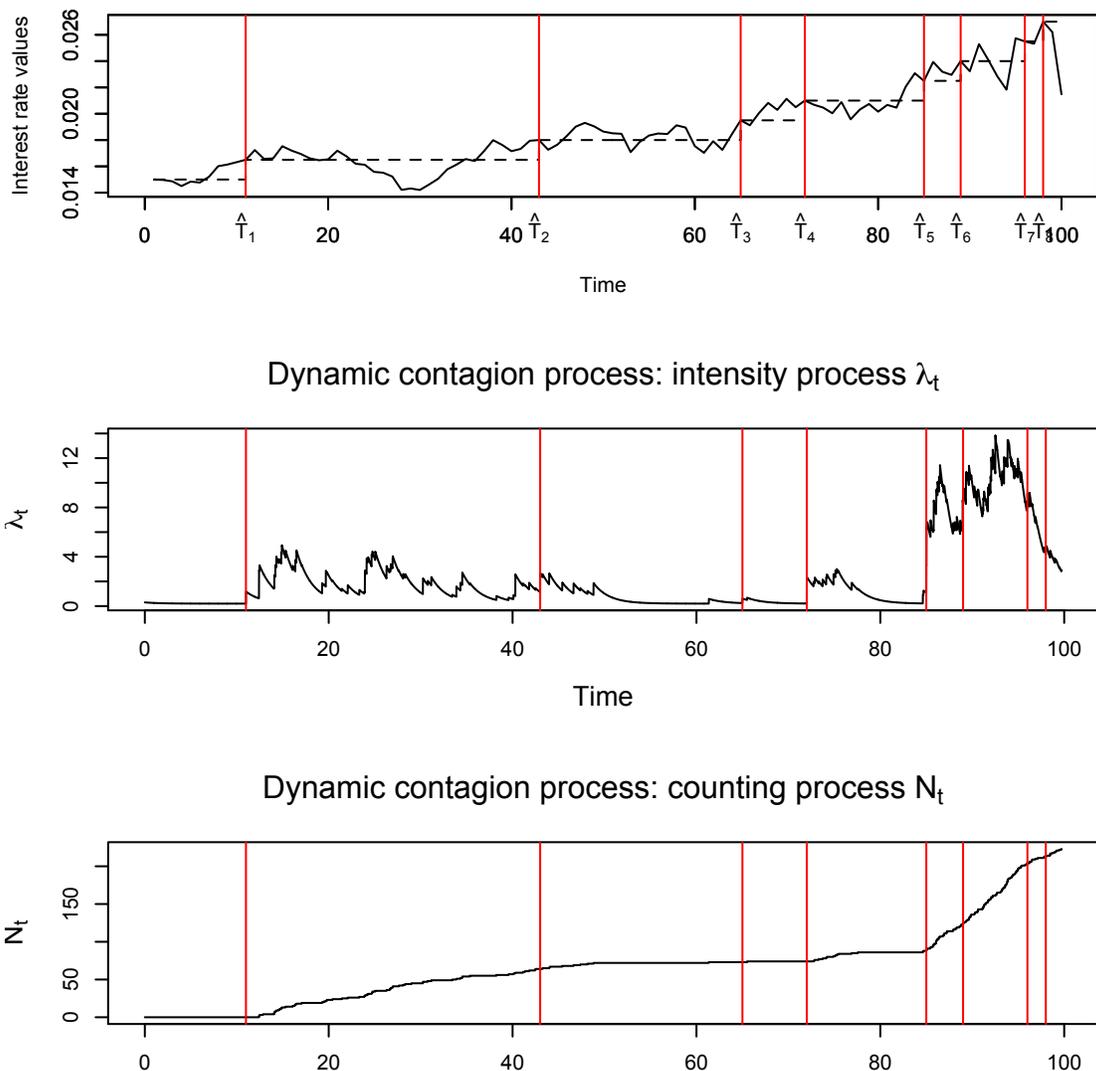


Figure 2: **Intensity dynamics  $\lambda_t$  and counting process  $N_t$ .** The two first graphs respectively show one particular sample path of the interest rate and the corresponding intensity  $\lambda_t$  given in Equation (5). Vertical lines stand for the external jump times. The third graph is associated to the counting process  $N_t$ . The barrier for the relative gap is  $B = 10\%$ , while the base case parameters of the dynamic contagion process are  $\lambda_0 = 0.5, \lambda_\infty = 0.4, \beta = 2, \delta = 2, \gamma = 1.5$ .

### 3.2 Recursive formulas for moments derivation

The aim of this section is to derive the moments of the intensity  $\lambda_t$  as well as of the number of lapses. To this end, we consider the *moment generating function* (MGF) of the intensity, denoted by  $m(t, \theta) = \mathbb{E}[e^{\theta\lambda_t}]$ . Let  $m^{(n)}(t, \theta)$  be the  $n^{\text{th}}$  derivative of  $m$  with respect to  $\theta$ , such that  $m^{(n)}(t, 0)$  is the  $n^{\text{th}}$  moment of  $\lambda_t$ . Seemingly, denote by  $\xi(t, \theta)$  and  $\hat{\xi}(t, \theta)$  the MGF of the processes  $Z_t$  and  $\hat{Z}_t$ , defined as

$$Z_t = \sum_{i=1}^{N_t} X_i e^{\beta T_i} \quad \text{and} \quad \hat{Z}_t = \sum_{i=1}^{\hat{N}_t} Y_i e^{\beta \hat{T}_i}. \quad (12)$$

These processes are discounted compounded renewal processes (see [15]), and  $\xi^{(n)}(t, \theta)$  and  $\hat{\xi}^{(n)}(t, \theta)$  refer to the  $n^{\text{th}}$  derivative of  $\xi(t, \theta)$  and  $\hat{\xi}(t, \theta)$  with respect to  $\theta$ . Hence, the process  $\lambda_t$  can be written in the following form

$$\lambda_t = (\lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t}) + e^{-\beta t} Z_t + e^{-\beta t} \hat{Z}_t,$$

which is useful for the results stated in the propositions below.

**Proposition 1** *The MGF of  $\lambda_t$  can be decomposed as follows:*

$$m(t, \theta) = e^{\theta(\lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t})} \xi(t, \theta e^{-\beta t}) \hat{\xi}(t, \theta e^{-\beta t}), \quad (13)$$

with  $\xi(t, \theta)$  and  $\hat{\xi}(t, \theta)$  being the MGF of  $Z_t$  and  $\hat{Z}_t$ .

**Proof:** see Appendix A.1 ■

The moments of the lapse intensity thus directly depend on those of  $Z_t$  and  $\hat{Z}_t$ , which are derived via a recursive formula ([15]).

**Lemma 1** *The MGFs  $\xi$ ,  $\hat{\xi}$  of  $Z_t = \sum_{i=1}^{N_t} X_i e^{\beta T_i}$  and  $\hat{Z}_t = \sum_{j=1}^{\hat{N}_t} Y_j e^{\beta \hat{T}_j}$  are given by the recursive formulas:*

$$\xi(t, \theta) = 1 + \int_0^t \left( \frac{\theta e^{\beta u}}{\gamma - \theta e^{\beta u}} \right) \xi(t - u, \theta e^{\beta u}) m^{(1)}(u, 0) du, \quad (14)$$

$$\hat{\xi}(t, \theta) = 1 + \int_0^t \left( \frac{\theta e^{\beta u}}{\delta - \theta e^{\beta u}} \right) \hat{\xi}(t - u, \theta e^{\beta u}) dh(u), \quad (15)$$

where  $h(t)$  is given in (11). The moments of  $Z_t$ ,  $\hat{Z}_t$  have the following form:

$$\xi^{(n)}(t, 0) = \sum_{k=0}^{n-1} \frac{n!}{k!} \frac{1}{\gamma^{n-k}} \int_0^t e^{n\beta u} \xi^{(k)}(t - u, 0) m^{(1)}(u, 0) du, \quad (16)$$

$$\hat{\xi}^{(n)}(t, 0) = \sum_{k=0}^{n-1} \frac{n!}{k!} \frac{1}{\delta^{n-k}} \int_0^t e^{n\beta u} \hat{\xi}^{(k)}(t - u, 0) dh(u), \quad (17)$$

**Proof:** see Appendix A.2. ■

Notice that the functions  $\xi^{(n)}(t, \theta)$  are mainly dependent on the knowledge of  $m^{(1)}(t, 0)$ , which is the expectation of the lapse intensity. This first moment can be inferred using arguments similar to those of [6], whose technique is based on the infinitesimal generator of a Markov process and a martingale argument. In their paper, the exponential distribution of the inter-arrival times of external jumps makes the process  $\lambda_t$  be Markovian, an assumption which is not fulfilled in our setting. Indeed, in our case, the lapse intensity  $\lambda_t$  is not a Markov process. Referring to the method in [9], the process  $\lambda_t$  can be transformed to a Markovian one by introducing a supplementary process, namely the time elapsed since the last external jump (or the time remaining to the next one). Here we rather rely on a differential argument based on the recursive formulas of Proposition 1 and Lemma 1. This allows us to deduce another recursive equation satisfied by the derivatives of  $m(t, \theta)$ .

**Proposition 2** *For  $n \geq 1$ , the  $n^{\text{th}}$  derivative of the lapse intensity MGF is given recursively as follows:*

$$\begin{aligned} m^{(n)}(t, \theta) &= (\lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t})m^{(n-1)}(t, \theta) \\ &\quad + \sum_{i=0}^{n-1} \binom{n-1}{i} e^{-(n-i)\beta t} \left( I_i(t, \theta) + \widehat{I}_i(t, \theta) \right) m^{(i)}(t, \theta), \end{aligned} \quad (18)$$

where  $I_k$  and  $J_k$  for  $\{k = 1, 2, \dots\}$  are given by

$$\begin{aligned} I_k(t, \theta) &= I_{k-1}^{(1)}(t, \theta) - kI_{k-1}(t, \theta)\xi^{(1)}(t, \theta e^{-\beta t}), \\ \widehat{I}_k(t, \theta) &= \widehat{I}_{k-1}^{(1)}(t, \theta) - k\widehat{I}_{k-1}(t, \theta)\widehat{\xi}^{(1)}(t, \theta e^{-\beta t}), \end{aligned}$$

with  $I_0(t, \theta) = \xi^{(1)}(t, \theta e^{-\beta t})/\xi(t, \theta e^{-\beta t})$  and  $\widehat{I}_0(t, \theta) = \widehat{\xi}^{(1)}(t, \theta e^{-\beta t})/\widehat{\xi}(t, \theta e^{-\beta t})$ .

**Proof:** see Appendix A.3. ■

A particular application of the above result is the derivation of the consecutive moments of the intensity process  $\lambda_t$  by letting  $\theta = 0$  in (18). This recursive formula is then used to get the analytic expression of the first moment of the intensity process as shown in the Lemma below.

**Lemma 2 (First moment)** *The expectation of the lapse intensity  $\lambda_t$  is given by*

$$\mathbb{E}[\lambda_t] := m^{(1)}(t, 0) = \left( \lambda_0 - \frac{\beta\lambda_\infty}{\beta - 1/\gamma} \right) e^{-(\beta - \frac{1}{\gamma})t} + \frac{\beta\lambda_\infty}{\beta - 1/\gamma} + \frac{1}{\delta} \int_0^t e^{-(\beta - \frac{1}{\gamma})(t-s)} h'(s) ds. \quad (19)$$

*It is the solution of the differential equation*

$$\frac{\partial m^{(1)}(t, 0)}{\partial t} = \beta\lambda_\infty - \left( \beta - \frac{1}{\gamma} \right) m^{(1)}(t, 0) + \frac{1}{\delta} h'(t),$$

with  $h(t)$  given in Equation (11).

**Proof:** see Appendix A.4. ■

This result is crucial for the derivation of some relevant risk indicators and the quantification of correlation and contagion effects on lapse rates. For instance, we can derive the expected number of lapses over a given time horizon. Using Lemma 2, we formally have:

$$\mathbb{E}[N_t] = \mathbb{E} \left[ \int_0^t \lambda_s ds \right] = \int_0^t m^{(1)}(s, 0) ds, \quad (20)$$

where Fubini-Tonelli's theorem has been used to intervene the expectation sign and the integral.

Notice that the first moment in Lemma 2 involves an infinite series associated with the external jumps component. Knowing that  $h(t) = \sum_{j=0}^{\infty} F^{j*}(t)$  is linked to a distribution function and has a density given by  $h'(t) = \sum_{j=0}^{\infty} f^{j*}(t)$  (where  $f^{j*}$  is the derivative of  $F^{j*}$ ), we can write

$$\mathbb{E}[\lambda_t] = \left( \lambda_0 - \frac{\beta \lambda_{\infty}}{\beta - 1/\gamma} \right) e^{-(\beta - \frac{1}{\gamma})t} + \frac{\beta \lambda_{\infty}}{\beta - 1/\gamma} + \frac{1}{\delta} \left( \sum_{j \geq 1} \int_0^t f^{j*}(s) e^{-(\beta - \frac{1}{\gamma})(t-s)} ds \right). \quad (21)$$

The numerical calculation of (21) is closely related to the computation of the infinite sum

$$\sum_{j \geq 1} \int_0^t f^{j*}(s) e^{-(\beta - 1/\gamma)(t-s)} ds.$$

Nevertheless, only the first  $k$  terms of this sum have a significant quantitative impact on the expectation, and the number of relevant terms  $k$  depends on the values of the parameters  $\theta_1$  and  $\theta_2$ . The inverse Gaussian density  $f^{j*}$  is flattening as  $j$  increases, which makes the product with the exponential density quickly go to zero. This also explains why there exists a closed-form formula for the expectation of the intensity process, as reported in Proposition 3. To illustrate this behavior, Figure 3 reports these densities for various values of  $j$ .

**Proposition 3** *Set  $k := (\theta_2/\theta_1^2) - 2(\beta - \frac{1}{\gamma})$ , where  $(\theta_1, \theta_2)$  are the parameters of the inverse Gaussian law given in (9). The first moment of the intensity process can be written in closed form as a function of the parameters defining the dynamic contagion process.*

(i) *If  $(\beta - \frac{1}{\gamma}) \leq \frac{(\mu - \sigma^2/2)^2}{2\sigma^2}$  (or equivalently  $k \geq 0$ ), then*

$$\begin{aligned} \mathbb{E}[\lambda_t] = & \left( \lambda_0 - \frac{\beta \lambda_{\infty}}{\beta - 1/\gamma} \right) e^{-(\beta - \frac{1}{\gamma})t} + \frac{\beta \lambda_{\infty}}{\beta - 1/\gamma} + \frac{1}{\delta} e^{-(\beta - \frac{1}{\gamma})t} \times \\ & \sum_{j \geq 1} e^{j \frac{\theta_2}{\theta_1}} \left[ e^{-j\sqrt{k}\theta_2} \Phi \left( \frac{t\sqrt{k} - j\sqrt{\theta_2}}{\sqrt{t}} \right) + e^{j\sqrt{k}\theta_2} \Phi \left( -\frac{t\sqrt{k} + j\sqrt{\theta_2}}{\sqrt{t}} \right) \right], \end{aligned} \quad (22)$$

(ii) *else*

$$\mathbb{E}[\lambda_t] = \left( \lambda_0 - \frac{\beta \lambda_{\infty}}{\beta - 1/\gamma} \right) e^{-(\beta - \frac{1}{\gamma})t} + \frac{\beta \lambda_{\infty}}{\beta - 1/\gamma} + \frac{1}{\delta} \sum_{j \geq 1} e^{-\frac{\theta_2}{\theta_1^2} \frac{(j\theta_1 - t)^2}{2t}} \times \text{Re}(w(z_j)), \quad (23)$$

where  $\text{Re}(\cdot)$  is the real part of  $w(z_j) = e^{-z_j^2} (1 - \text{erf}(-iz_j))$ , with  $z_j = \sqrt{\frac{-kt}{2}} + ij\sqrt{\frac{\theta_2}{2t}}$  and  $\text{erf}(z) = (2/\sqrt{\pi}) \int_0^z e^{-t^2} dt$ .

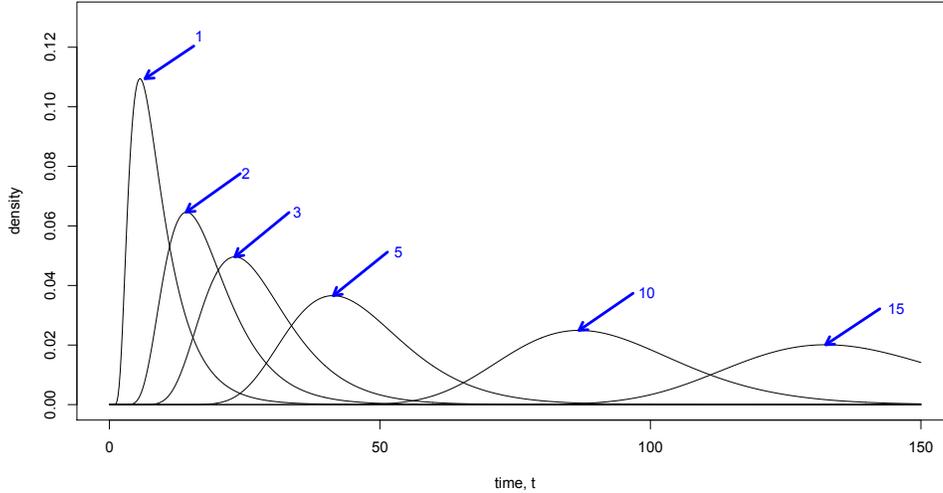


Figure 3: **Density  $f^{j*}(t)$  of the sum of  $j$  i.i.d. inverse Gaussian random variables.** The graph reports the behavior of the density function  $f^{j*}(t)$  of an inverse Gaussian law for different values of  $j \in \{1, 2, 3, 5, 10, 15\}$ , where the density function  $f(t)$  is given in Equation (9).

**Proof:** see Appendix A.5. ■

We are now interested in deriving the limiting behavior (as  $t$  tends to infinity) of the mean intensity, since it is a key driver for monitoring the lapse risk profile. This can be done by leveraging on the stability assumption of the form  $\gamma\beta > 1$ , similar to the one used in the original Hawkes framework (see [2] for general decay functions, as well as the externally excited framework considered in [6] and [1] among others). This condition ensures the stability of the lapse intensity in the long run, i.e. the lapse intensity converges to its long-run equilibrium level. The following proposition reports the limiting behavior of the first moment of the intensity process  $\lambda_t$ .

**Proposition 4** *The limiting expected lapse intensity can be written as follows:*

$$\lim_{t \rightarrow \infty} \mathbb{E}[\lambda_t] = \frac{\beta\lambda_\infty}{\beta - 1/\gamma} + \frac{1}{\delta\theta_1(\beta - 1/\gamma)}. \quad (24)$$

with  $\mathbb{E}[\lambda_t]$  given in Equation (21).

**Proof:** see Appendix A.6. ■

## 4 Lapse intensity: sensitivity to the model parameters

Our goal is to understand the role that each model parameter plays in our framework, in terms of influence on the lapse risk. To measure it from both a qualitative and quantitative viewpoint, we consider the expectation of the intensity process  $\mathbb{E}[\lambda_t]$  as the main risk indicator, see Proposition 3 and Figure 4.

This indicator has been chosen for two reasons. First, it gives the most relevant information concerning the impact of the dynamic integration of correlation and contagion on the mean risk level supported by the insurance company. Secondly, it enables to derive other qualitative and quantitative results (such as the average number of lapses) quite easily. The sensitivity analysis presented below is based on a *ceteris paribus* comparison of the proposed dynamic contagion process  $\lambda_t$  in (21), whose base case parameters are given in Figure 4 (left side).

To start with, let us consider how changes of the GBM parameters affect the first moment of the intensity process  $\lambda_t$ . First, look at the drift  $\mu$ : according to Figure 5, the greater the drift the higher the mean intensity process (for each time horizon). A greater drift naturally makes the exogenous barrier  $B$  be hit more often, thus increasing the external risk component by generating the occurrence of additional external jumps. The resulting long-term mean of the intensity process is also higher, while the time to reach this stationary equilibrium regime does not seem to be significantly impacted. Figure 5 shows that this long-term behavior is approximately reached at  $t = 50$ , for each considered level of the drift parameter  $\mu$ . On the contrary, the volatility  $\sigma$  reveals to have an impact on the time needed by the mean intensity to reach this long-run equilibrium: the more  $\sigma$  increases, the longer this time. Increasing the volatility in the interest rate dynamics means increasing the level of uncertainty characterizing the system, and more uncertainty obviously postpones the stability. In mathematical terms, this can be seen by looking at Equation (24) and Proposition 3: increasing  $\sigma$  makes the mean intensity process raising in the short run and decreasing in the long-term.

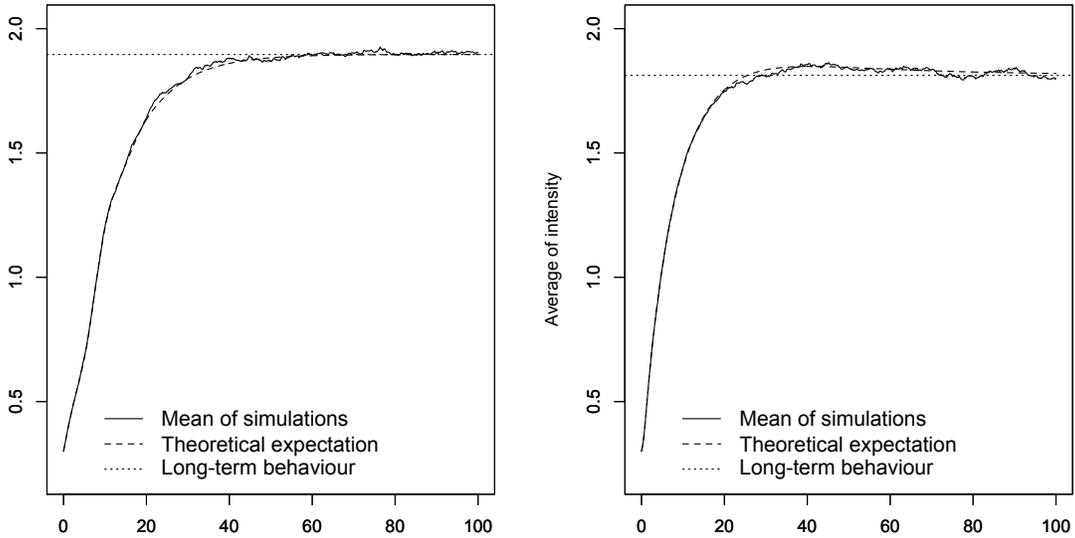


Figure 4: **Expectation of the lapse intensity process for  $t \in [0, 100]$ .** The plots compare the outcome of a Monte Carlo simulation ( $2 \cdot 10^4$  trajectories) to the theoretical expressions of Proposition 3. On the *left* the case of Equation (22), with parameters:  $r_0 = 1.5\%$ ,  $\mu = 1\%$ ,  $\sigma = 1\%$ ,  $B = 10\%$ ,  $\lambda_0 = 0.3$ ,  $\lambda_\infty = 0.2$ ,  $\delta = 1.5$ ,  $\gamma = 2$ , and  $\beta = 0.6$ . The plot on the *right* corresponds to the case of Equation (23), with the same values of parameters except that  $\sigma = 5\%$ . The two plots also report the long term value of the lapse intensity derived in Equation (24).

Now, focus on the exogenous barrier  $B$  defined for the relative gaps  $RG_t^j$  capturing the spread between the contract credited rate and the market rate. The barrier  $B$  can be seen as a proxy for an elasticity-based measure in a pure economic sense: this threshold is indeed introduced to reflect the policyholders' sensitivity to interest rates movements. It is therefore linked to the propensity of policyholders to react when the relative gap tends to increase: the higher  $B$  gets, the lower the mean intensity process is. More external jumps will occur if  $B$  is low, making  $\lambda_t$  greater most of the time. Though, the long-run stable regime does not change when we modify this barrier:  $B$  is indeed not appearing in (24) while it is directly impacting the external source of risk affecting  $\lambda_t$ .

Let us consider now the parameters of the intensity process itself: the exponential decay  $\beta$  has one of the most relevant impacts. This parameter represents the speed at which the intensity  $\lambda_t$  decreases once a jump has occurred. Not surprisingly, a low  $\beta$  has two main consequences: i) it makes the time to reach the long-run expected level longer; ii) it considerably increases the mean intensity  $\mathbb{E}[\lambda_t]$ , for each date  $t$ . As time passes, a low  $\beta$  increases the gap with respect to the base case intensity process, highlighting that the effect of  $\beta$  clearly accumulates over time and leads to a much higher long-term mean intensity.

The same argument applies when considering a variation of  $\gamma$  parameter, which is in line with the stationarity condition: as long as  $\beta$  is greater than the mean size  $1/\gamma$  of self-excited jumps, the mean intensity process remains finite in the long run. This relation explains why their effect is similar: when  $\gamma$  gets higher, the mean size of internal jumps decreases, so that the mean lapse intensity process is lower, as well as the elapsed time before reaching the stationary level. Recall that these two parameters capture the self-excited component of our dynamic contagion process.

Finally, changing the parameter  $\delta$  produces an impact on the external component of the intensity process. This impact is similar to the influence produced by a bump in the GBM drift  $\mu$  and the exogenous barrier  $B$ , since the ratio  $1/\delta$  directly relates to the mean size of external jumps. The initial intensity level  $\lambda_0$  has no long-term effect on the mean intensity level and on the time to reach the stationary regime; whereas  $\lambda_\infty$  obviously impacts the long-term mean intensity. This is consistent with the analytic expression given in (21), where it is evident that  $\mathbb{E}[\lambda_t]$  increases if  $\lambda_\infty$  is higher. Table 1 summarizes all these results and Figure 5 gives a graphical illustration of the qualitative behavior described above.

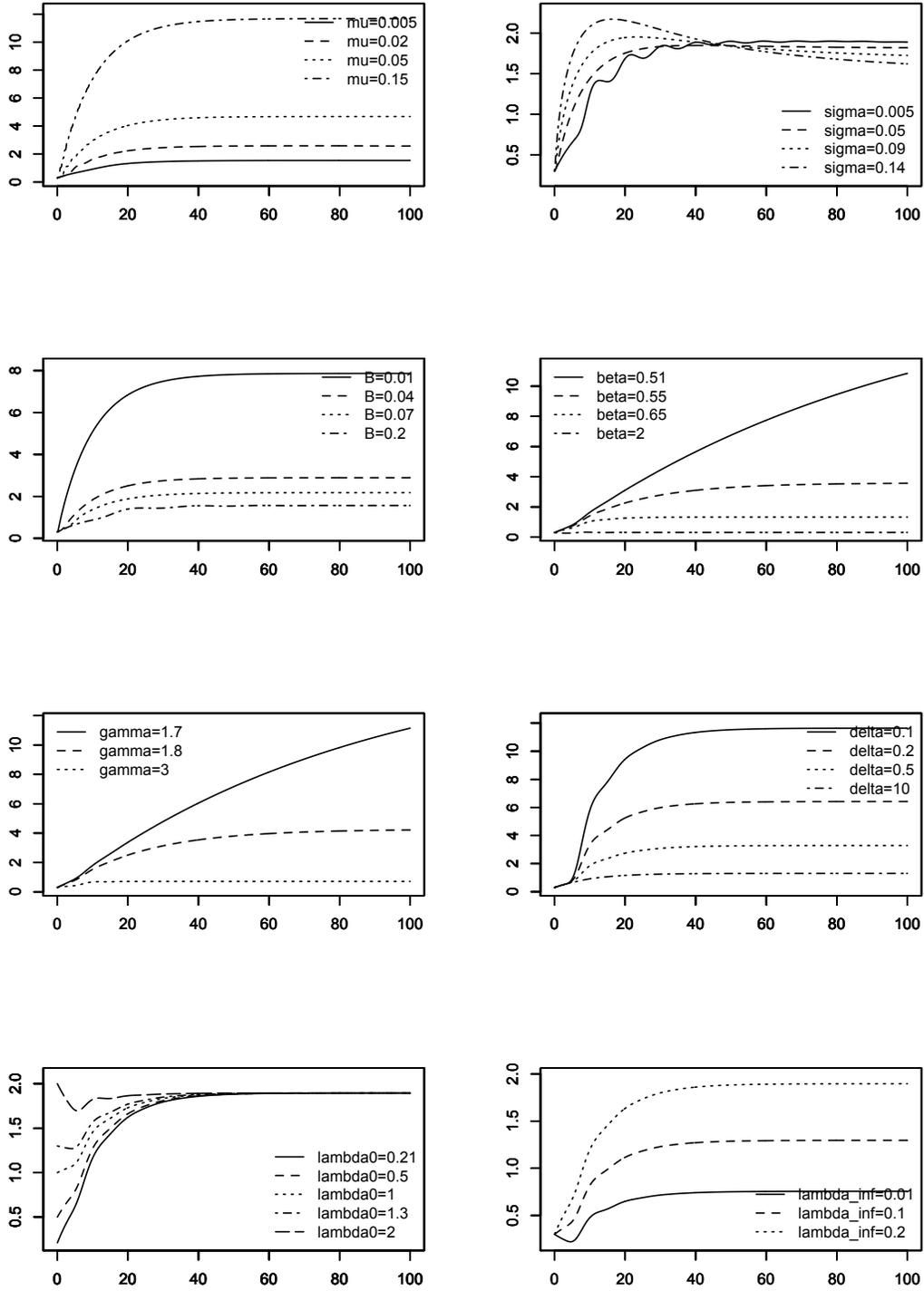


Figure 5: Sensitivity of the expected lapse intensity to the model parameters for  $t \in [0, 100]$ . The comparison is made referring to the case depicted on the left in Figure 4, with the same parameter values. This is a coeteris paribus analysis: everything else fixed, the sensitivity is analyzed by changing only the value of the parameter under analysis, i.e. only one parameter each time.

Reference model	Profile of trajectory	Mean intensity level $\mathbb{E}[\lambda_t]$ :			Time to reach stationarity
		short-term	mid-term	long-term	
	<i>Figure 4 (left)</i>				
$\mu \uparrow$	unchanged	+	+	+	$\approx$
$\sigma \uparrow$	modified	+	$\approx$	-	+
$B \uparrow$	unchanged	-	-	-	$\approx$
$\beta \uparrow$	modified	-	-	-	-
$\gamma \uparrow$	modified	-	-	-	-
$\delta \uparrow$	unchanged	-	-	-	-
$\lambda_0 \uparrow$	modified	+	$\approx$	=	$\approx$
$\lambda_\infty \uparrow$	unchanged	+	+	+	$\approx$

Table 1: **Qualitative impact of the model parameters on the expected lapse intensity.** We summarize the impact of an increase in each model parameter on  $\mathbb{E}[\lambda_t]$ . The sign “+” corresponds to a positive impact (increase) on the quantity of interest (as compared to the reference case, see Figure 4 left side); the sign “-” stands for a decrease of  $\mathbb{E}[\lambda_t]$ ; and “ $\approx$ ” means that there is roughly no change, i.e. the impact is quite negligible.

## 5 Application to risk management

From a risk management perspective, insurance companies are interested in properly capturing and monitoring the lapse risk (linked to underlying surrender options) for hedging and capital reserving purposes. Thus, estimating key measures driving this source of risk becomes a crucial activity to be carried out on a regular basis. In this section we perform a comparative study between the most widely used approaches currently considered to estimate potential losses in practice (S-shaped curve and standard formula of Solvency II) and the outcomes of our methodology. The study is conducted via a simulation exercise, useful to illustrate the main ideas of the paper and their implications in terms of real life practical issues for a company’s risk profile estimation.

The first step of the analysis is the calibration of the parameters involved in the proposed model, and the question is: *How can the model be calibrated on real data by an insurance company?* Consider firstly the initial intensity level  $\lambda_0$  and its long-run stationary mean  $\lambda_\infty$ . As already discussed in Section 2,  $\lambda_0$  represents the initial force of lapse and is a constant: the underlying lifetime distribution before lapsation is thus exponential. Therefore, for a given insurance portfolio, a suitable value for  $\lambda_0$  could be the inverse of the *empirical mean of times before lapsation* in a stable regime. Inside insurance companies, this quantity is usually well-known, and an *a priori* segmentation could be performed to integrate some risk factors specific to the structural component underlying lapses. Assume that there are 1.000 contracts in the portfolio, and that the company stores data on a daily basis. Up to the date of the analysis, let us say that 3 lapses are observed every 10 trading days, which means that  $\lambda_0 = 0.3$ . In a 1-year time horizon (250 trading days), the company should experience around 75 lapses (hence a yearly structural lapse rate of 7.5%). With a similar logic, the long-run lapse intensity level  $\lambda_\infty$  can be fixed by the risk managers as the target level embedding limits coming from risk management purposes. Indeed, when the time horizon is given, this quantity is linked to the final expected number of lapses since the mean contract lifetime would approximately be known (at least under the assumption of a stable economic

regime). Say for instance that the insurer would like to experience no more than 2 lapses every 10 trading days, or equivalently 50 structural lapses a year. The mean inter-arrival time for these events is thus 5 days, which leads to  $\lambda_\infty = 0.2$ .

Other parameters ideally have to be calibrated from empirical data, still keeping in mind that some of them are connected to the risk appetite of the company. For instance,  $\beta$  relates to the ability of the company to reassure the policyholders, and parameters  $\gamma$  and  $\delta$  tie in with the mean size (in terms of sum insured) of lapsed contracts. Typically,  $\delta$  should be lower than  $\gamma$ , because the lapsation by the richest policyholders may have a deeper (even much more critical) impact on the insurer's balance sheet. Since they are often advised, the richest individuals globally behave more rationally, and are the first to react in unfavorable situations. In other words, external jumps should be fewer but have a more significant quantitative impact on the  $P\&L$  of the company. The parameters of the interest rate dynamics can be calibrated from historical data, while the barrier  $B$ , capturing the sensitivity of policyholders to market movements, usually comes from an expert-based estimation (e.g. depending on companies, it can vary between 10% and 50%).

We now compare the different methodologies (current market practices and ours) to provide some key risk indicators, in particular the expected number of lapses and associated risk measures. To be compliant with the Solvency II framework, we work under the following assumptions: i) the insurance portfolio is in run-off, and ii) the reference time horizon is 1 year. With respect to the current financial context characterized by extremely low interest rates, we focus on potential parallel upward shocks of interest rates curves and the corresponding potential massive lapses (i.e. the most likely problematic scenario). The risk measures under consideration are the *Value-at-Risk* ( $VaR$ ) and the *Tail-Value-at-Risk* ( $TVaR$ ), respectively defined at the threshold  $\alpha \in (0, 1)$  by

$$VaR_\alpha(N_t) = \inf\{k : \mathbb{P}(N_t \leq k) \geq \alpha\}; \quad TVaR_\alpha(N_t) = \frac{1}{1-\alpha} \int_\alpha^1 VaR_p(N_t) dp, \quad (25)$$

with  $k \in \mathbb{N}$ : they enable to quantify the riskiness of behaviors' changes via the tail of the distribution of the number of lapses. Indeed,  $VaR_\alpha(N_t)$  tells us that the number  $N_t$  of lapses for a given period  $t$  will be lower than  $VaR_\alpha(N_t)$  with probability  $\alpha$ . It somewhat gives the information about how the SCR would be modified by the consideration of both correlation and contagion effects. The  $TVaR_\alpha$  risk measure is nothing else than the average of  $VaR_p$  from the threshold  $\alpha$  to 1, and thus informs us about the whole behavior of the tail.

Numerical results are summarized in Table 2. The first conclusion is straightforward: looking at the results about shocks and  $VaR_\alpha(N_t)$ , it seems that classical stress tests procedures (e.g. Solvency II and S-shaped approaches) lead to an overall underestimation of the *actual* lapse risk faced by life insurance companies. This is not surprising, since correlation and contagion effects among policyholders behaviors are not taken into account in such methodologies. We should also mention that the results obtained from the Solvency II standard formula are neither really consistent nor realistic here, since the Solvency II shock applied to the baseline risk was calibrated on a single product in a specific market. In practice, the resulting estimation is thus not conservative and a greater reserve (roughly 2.5 times bigger than using a S-shaped curve approach, even more in the Solvency II case) should instead be considered to cover unexpected potential losses coming from lapses. Concerning the mean risk level as well as stressed scenarios, notice that the standard Hawkes process ([11]) provides lower bounds for the dynamic contagion case, which is consistent with their economic and financial interpretation. Moreover, the  $TVaR_\alpha(N_t)$  results show that the

Parameters	Solvency II Standard formula		S-shaped curve (ONC)		Hawkes counting process			Dynamic contagion process			
	Risk level	Shocks	Risk level	Shocks	$\mathbb{E}[N_t]$	$VaR_\alpha(N_t)$	$TVaR_\alpha(N_t)$	$\mathbb{E}[N_t]$	$VaR_\alpha(N_t)$	$TVaR_\alpha(N_t)$	
$B$	10%							455	1028	1142	
	30%	75	112 <sup>7</sup>	75	375 <sup>8</sup>	291	776	837	312	818	930
	50%								293	778	886
$\delta$	0.1							2461	4286	4559	
	0.5	75	112	75	375	291	776	837	702	1460	1594
	1.5								455	1028	1142

Table 2: **Impact of contagion and correlation on risk metrics for  $N_t$ .** Various levels of the barrier  $B$  and mean external jump size  $1/\delta$  are considered to study the results on  $VaR_\alpha(N_t)$  and  $TVaR_\alpha(N_t)$ , with  $\alpha = 99.5\%$  and  $t = 250$  (1-year time horizon). All other model parameters remain the same as in Figure 4 (left).

right tail of the distribution of the number of lapses is heavier in our setting than in all other cases, including the Hawkes model.

Let us now focus on the policyholders' sensitivity to interest rate movements. Table 2 highlights a non-linear behavior. Starting from an exogenous barrier  $B$  equal to 50%, a 20%-decrease of  $B$  causes a 7%-increase of the expected number of lapses, whereas the latter rises by around 30% when  $B$  goes from 30% to 10%. This is confirmed by Figure 6, which considers the relationship between the Value-at-Risk of the number of lapses obtained through our methodology based on dynamic contagion, namely  $VaR_\alpha^{Dc}(N_t)$ , and the one obtained by following Solvency II guidelines, namely  $VaR_\alpha^{Sol}(N_t)$ . The idea is to study the functional relationship between the Value-at-Risk of the number of lapses under the two approaches and analyze how this relationship (e.g. shape and magnitude of the ratio  $VaR_\alpha^{Dc}(N_t)/VaR_\alpha^{Sol}(N_t)$ ) is affected by: i) the barrier  $B$  for the interest rate relative gaps, determining the times at which external jumps are observed; ii) the mean size of external jumps, i.e. through the variation of  $\delta$  parameter.

The  $z$ -axis in Figure 6 expresses the ratio of these two Value-at-Risk measures, by reporting  $VaR_\alpha^{Dc}(N_t)/VaR_\alpha^{Sol}(N_t)$ . Most of time, the introduction of correlation through the external component seems to have a reduced but not negligible impact on the quantities of interest. As we can see from Figure 6, while the mean size of external jumps remains lower than a certain threshold (roughly given by  $\delta = 0.4$ ), the risk measures tend to increase slowly (yellow area). However, as soon as the mean size of external jumps increases, the relationship between both quantities becomes instead exponential (orange and red areas), highlighting the material impact of this component. The economic insight behind this behavior suggests that the portfolio composition is crucial, and insurers most exposed to lapse risk are the ones having a great proportion of rich policyholders. Indeed, this population is usually the first-to-react, with possibly huge amount of notional associated to the surrenders (a big notional is associated to a low  $\delta$ ). The proportionality coefficient between those two VaRs ( $VaR_\alpha^{Dc}(N_t)/VaR_\alpha^{Sol}(N_t)$ ) can be very high. For instance it reaches 40 in our example, when the mean size of external jumps is set to 5 (e.g.  $\delta = 0.2$ ). Remember that these results are only illustrative in the sense that the parameters were not calibrated on real his-

<sup>7</sup>Defined in the calibration papers of the 5<sup>th</sup> QIS: up to 1.5 times the structural lapses.

<sup>8</sup>Defined in the ONC (see Section 1): corresponds to structural lapses to which temporary lapses are added (with a maximum of 30% of the global exposure). Here it is thus  $75 + 30\% \times 1000 = 375$  lapses.

torical data, but they corroborate the idea that standard stress-test procedures are not capturing a significant component underlying the *real* lapse risk faced by life insurance companies.

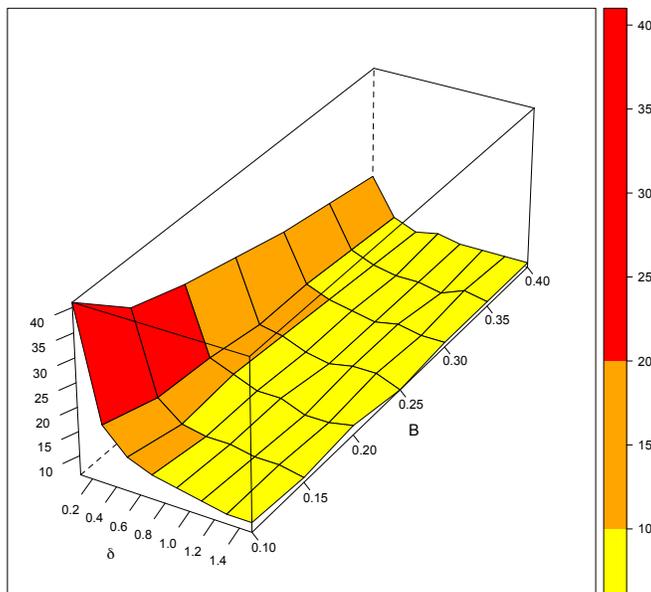


Figure 6: **Impact of  $B$  and the mean size of external jumps  $1/\delta$  on  $VaR_\alpha(N_t)$ .** The z-axis expresses the ratio between the dynamic contagion-based risk metric and the one based on Solvency II definition. This relationship is studied as function of two drivers: the barrier  $B$  defined for the interest rates relative gap; and the mean size of external jumps through the parameter  $\delta$ .

## 6 Conclusion

This paper defines a methodology that enables to deal with both stressed and unstressed scenarios for the risk management activity of surrender events. Due to the optionality embedded in most life insurance contracts, correlation and contagion among policyholders' behaviors in reaction to interest rates movements appear to be crucial drivers underlying a proper measurement of the lapse risk. We show through numerical examples that by properly considering their impact, we could end up to around a 2.5 factor times the capital requirement obtained via standard S-shaped/Solvency II procedures. This is the reason why insurance companies must pay attention to correlation and contagion when evaluating both risk margins and related capital requirements: a proper measurement is fundamental for capital reserving purposes, for the definition of adequate stress test scenarios and for the building up of proper stress testing procedures. We propose a mathematical framework that allows to dynamically integrate these effects into the modeling of the lapse risk (*dynamic contagion model*) and give in closed form the moments of the lapse intensity process. Our numerical analysis highlights that the lapse risk margins are probably underestimated when using current practices (e.g. Solvency II-based) and the magnitude of this underestimation can be significant in terms of capital reserves depending on the sensitivity of policyholders' reactions to changes in the financial

environment (i.e. *timing risk* for the surrender event, number of surrenders, notional involved).

Further developments of the model could focus on different aspects: separating the impacts of a global crisis scenario affecting all policyholders and a critical scenario only affecting some of them; alternative definitions of the barrier  $B$  embedding time dependency, in order to capture a dynamic sensitivity of policyholders. In the same spirit, we could also think of introducing delayed reactions of both policyholders and insurance companies, respectively to surrender the contract and to adjust the level of the guaranteed rates. However, these extensions go beyond the scope of this paper and we leave the investigation of such interesting questions for future research.

## References

- [1] Boumezoued, A., *Population viewpoint on Hawkes processes*, Working paper, arXiv preprint arXiv:1504.06563, (2015).
- [2] Brémaud, P. and Massoulié, L., *Power Spectra of General Shot Noises and Hawkes Point Processes with a Random Excitation*, Advances in Applied Probability, Vol. 34:1 (2002), p. 205–222.
- [3] Buchardt, K. and Moller, T. and Schmidt, K. B., *Cash Flows and Policyholder Behaviour in the semi-Markov life insurance setup*, Scandinavian Actuarial Journal, Vol. 8 (2015), p.1–29.
- [4] Dassios, A. and Jang, J.-W., *Pricing of Catastrophe Reinsurance and Derivatives Using the Cox Process with Shot Noise Intensity*, Finance & Stochastics, Vol. 7:1 (2003), p. 73–95.
- [5] Dassios, A. and Jang, J.-W., *A bivariate shot noise self-exciting process for insurance*, Insurance: Mathematics and Economics, Vol. 53:3 (2013), p. 524–532.
- [6] Dassios, A. and Zhao, H., *A dynamic contagion process*, Advances in Applied Probability, Vol. 43:3, (2011), p. 814–846.
- [7] Davis, M.H.A., *Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models*, Journal of the Royal Statistical Society. Series B. Methodological, Vol. 46:3 (1984), p. 353–388.
- [8] Dutang, C., *Regression models of price elasticity in non-life insurance*, Mémoire confidentiel - AXA Group Risk Management, (2011).
- [9] Embrechts, P. and Schmidli, H. and Grandell, J., *Finite-time Lundberg inequalities in the Cox case*, Scandinavian Actuarial Journal, n. 1 (1993), p. 17–41.
- [10] Errais, E. and Giesecke, K. and Goldberg, L.R., *Affine Point Processes and Portfolio Credit Risk*, Siam Journal of Financial Mathematics, Vol. 1 (2010), p. 642–665.
- [11] Hawkes, A.G. and Oakes, D., *A cluster process representation of a self-exciting process*, Journal of Applied Probability, (1974), p. 493–503.
- [12] Kagraoka, Y., *Modeling Insurance Surrenders by the Negative Binomial Model*, Working Paper, (2005) Musashi University (Japan)

- [13] Kim, C., *Modeling Surrender and Lapse Rates with Economic Variables*, North American Actuarial Journal, (2005), p. 56–70.
- [14] Kuo, W., Tsai, C., and Chen, W.-T., *An Empirical Study on the Lapse Rate: the Cointegration Approach*, Journal of Risk and Insurance, 70:3 (2003), p. 489–508.
- [15] Léveillé, G. and Garrido, J. and Fang Wang, Y., *Moment generating functions of compound renewal sums with discounted claims*, Scandinavian Actuarial Journal, n. 3 (2010), p. 165–184.
- [16] Loisel, S. and Milhaud, X., *From deterministic to stochastic surrender risk models: impact of correlation crises on economic capital*, European Journal of Operational Research, Vol. 214:2 (2011), p. 348–357.
- [17] Loisel, S. and Arnal, P. and Durand, R., *Correlation crises in insurance and finance, and the need for dynamic risk maps in ORSA*, Working paper (2010), hal-00502848.
- [18] Milhaud, X., *Exogenous and endogenous risk factors management to predict surrender behaviours*, ASTIN Bulletin, Vol. 43:3 (2013), p. 373–398.
- [19] Outreville, J.-F. , *Whole-life insurance lapse rates and the emergency fund hypothesis*, Insurance: Mathematics and Economics, Vol. 9 (1990), p. 249–255.
- [20] Renshaw, A. E. and Haberman, S., *Statistical analysis of life assurance lapses*, Journal of the Institute of Actuaries, Vol. 113 (1986), p. 459–497.
- [21] Schwarz, W., *On the convolution of inverse gaussian and exponential random variables*, Communications in Statistics - Theory and Methods, Vol. 31:12 (2002), p. 2113–2121.
- [22] Serfozo, R., *Renewal and Regenerative Processes*, In *Basics of Applied Stochastic Processes*, Springer (2009), p. 99–167.
- [23] Seshadri, V., *The inverse Gaussian distribution*, Springer (1999).
- [24] Shreve, S.E. and Karatzas, I., *Brownian motion and stochastic calculus*, Springer (1991).

## A Proofs

This Appendix contains the proofs of the main theoretical results derived in this paper.

### A.1 Proof of Proposition 1

The function  $m$  can be written as

$$m(t, \theta) = \mathbb{E} \left[ e^{\theta(\lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t})} e^{\theta e^{-\beta t} Z_t} e^{\theta e^{-\beta t} \widehat{Z}_t} \right].$$

Then, the decomposition follows by conditioning on  $I = \{T_1, \dots, T_{N_t+1}\} \cup \{X_1, \dots, X_{N_t+1}\}$ :

$$\begin{aligned} m(t, \theta) &= e^{\theta(\lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t})} \mathbb{E} \left[ \mathbb{E} \left[ e^{\theta e^{-\beta t} \widehat{Z}_t} \middle| I \right] e^{\theta e^{-\beta t} Z_t} \right], \\ &= e^{\theta(\lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t})} \xi(t, \theta e^{-\beta t}) \widehat{\xi}(t, \theta e^{-\beta t}), \end{aligned}$$

since  $\widehat{Z}_t$  is independent of  $I$  in the second equality. ■

## A.2 Proof of Lemma 1

(i) The first identities directly follow from [15] as:

$$\xi^{(n)}(t, 0) = \sum_{k=0}^{n-1} \frac{n!}{k!} \frac{1}{\gamma^{n-k}} \int_0^t e^{n\beta u} \xi^{(k)}(t-u, 0) dh(u),$$

where  $h(t) = \mathbb{E}[N_t]$  is the renewal function. Note that due to the martingale property of  $N_t - \int_0^t \lambda_u du$  and Fubini-Tonelli's theorem we can write  $d\mathbb{E}[h(u)] = \mathbb{E}[\lambda_u] du$ . Substituting this in the last equation and noting that  $\mathbb{E}[\lambda_t] = m^{(1)}(t, 0)$  leads to the desired formula.

(ii) The equalities directly follow from the results in [15] up to some basic algebraic calculation. ■

## A.3 Proof of Proposition 2

Let  $I_0(t, \theta) = \xi^{(1)}(t, \theta e^{-\beta t}) / \xi(t, \theta e^{-\beta t})$ , and consider the decomposition in (13) and the consecutive derivation. First, we have

$$\begin{aligned} m^{(1)}(t, \theta) &= \left( (\lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t}) + e^{-\beta t} \frac{\xi^{(1)}(t, \theta e^{-\beta t})}{\xi(t, \theta e^{-\beta t})} + e^{-\beta t} \frac{\widehat{\xi}^{(1)}(t, \theta e^{-\beta t})}{\widehat{\xi}(t, \theta e^{-\beta t})} \right) m(t, \theta), \\ &= (\lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t}) m(t, \theta) + e^{-\beta t} \left( I_0(t, \theta) + \widehat{I}_0(t, \theta) \right) m(t, \theta). \end{aligned}$$

Secondly, we get by differentiation of the above equation with respect to  $\theta$  the function  $m^{(n)}(t, \theta)$

$$\begin{aligned} m^{(2)}(t, \theta) &= (\lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t}) m^{(1)}(t, \theta) + e^{-\beta t} \left( I_0(t, \theta) + \widehat{I}_0(t, \theta) \right) m^{(1)}(t, \theta) \\ &\quad + e^{-\beta t} \left( I_1^{(1)}(t, \theta) + \widehat{I}_1^{(1)}(t, \theta) \right) m^{(1)}(t, \theta). \end{aligned}$$

Thus, by induction, some algebraic calculus and the relationship linking the  $I_k(t, \theta)$ 's and (resp.  $\widehat{I}_k(t, \theta)$ ) we come with the desired recursion of the  $m^{(n)}(t, \theta)$ . ■

## A.4 Proof of Lemma 2

(i) Letting  $n = 1$  in Equation (18), gives

$$m^{(1)}(t, \theta) = \left( (\lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t}) + e^{-\beta t} \frac{\xi^{(1)}(t, \theta e^{-\beta t})}{\xi(t, \theta e^{-\beta t})} + e^{-\beta t} \frac{\widehat{\xi}^{(1)}(t, \theta e^{-\beta t})}{\widehat{\xi}(t, \theta e^{-\beta t})} \right) m(t, \theta).$$

Now, letting  $\theta = 0$  and noting that  $m(t, 0) = \xi(t, 0) = \widehat{\xi}(t, 0) = 1$ , we can write the above equation as follows

$$m^{(1)}(t, 0) = (\lambda_\infty + (\lambda_0 - \lambda_\infty)e^{-\beta t}) + e^{-\beta t} \xi^{(1)}(t, 0) + e^{-\beta t} \widehat{\xi}^{(1)}(t, 0). \quad (26)$$

This equation does not directly yield to the desired formula for  $m^{(1)}(t, 0)$ , as the  $\xi^{(1)}(t, 0)$  depends on  $m^{(1)}(t, 0)$ . Indeed, we have from (16)

$$\xi^{(1)}(t, 0) = \frac{1}{\gamma} \int_0^t e^{\beta u} m^{(1)}(u, 0) du,$$

which makes (26) be an integral equation. Thus, it suffices to substitute  $\xi^{(1)}(t, 0)$  and  $\widehat{\xi}^{(1)}(t, 0)$  (given in (17)) into (26) and take the derivative with respect to the first variable to get the desired differential equation.

(ii) The explicit form is straightforward given the initial condition  $m^{(1)}(0, 0) = \lambda_0$ . ■

### A.5 Proof of Proposition 3

Observe that the integral in the last term of Equation (21) can be interpreted as the convolution of inverse Gaussian and exponential random variables, see [21]. By defining  $k := (\theta_2/\theta_1^2) - 2(\beta - \frac{1}{\gamma})$ , where  $(\theta_1, \theta_2)$  are the parameters of the inverse Gaussian law, we can directly derive the above result for the first moment of the intensity process by analytic computation. ■

### A.6 Proof of Proposition 4

The result follows by applying the stability condition  $\gamma\beta > 1$ . Observe that the limiting behavior of the lapse intensity first moment strongly depends on the limit of the last term in (19). The latter is nothing but the convolution of the derivative  $h'$  of the renewal function with the exponential function, with  $h(t)$  given in (11). Using the key renewal theorem, see [22, Theorem 35, p. 119], we have

$$\lim_{t \rightarrow \infty} \int_0^t e^{-(\beta - \frac{1}{\gamma})(t-s)} dh(s) = \frac{1}{\theta_1} \int_0^\infty e^{-(\beta - \frac{1}{\gamma})(t-s)} ds = \frac{1}{\theta_1} \frac{1}{\beta - 1/\gamma}.$$

As a consequence, the first moment given in Equation (21) tends to the limit given in (24) as  $t$  goes to  $+\infty$ . ■