



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. Insurance: Mathematics and Economics, 2016, 71, pp.317-331. 10.1016/j.insmatheco.2016.09.008 . hal-01282601v2

HAL Id: hal-01282601

<https://hal.science/hal-01282601v2>

Submitted on 15 Sep 2016

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* Xavier Milhaud^{†(2)} Yahia Salhi^{†(2)}

⁽¹⁾Risk Methodologies, UniCredit Spa, Milan, Italy

⁽²⁾Univ Lyon, ISFA, LSAF EA2429, F-69007 Lyon, France

September 15, 2016

Abstract

The present paper proposes a new methodology to model the lapse risk in life insurance by integrating the dynamic aspects of policyholders' behaviors and the dependency of the lapse intensity on macroeconomic conditions. Our approach, suitable to stable economic regimes as well as stress scenarios, introduces a mathematical framework where the lapse intensity follows a dynamic contagion process, see [11]. This allows to capture both contagion and correlation potentially arising among insureds' behaviors. In this framework, an *external market driven* jump component drives the lapse intensity process depending on the *interest rate trajectory*: when the spread between the market interest rates and the contractual crediting rate crosses a given threshold, the insurer is likely to experience more surrenders. A *log-normal dynamic* for the *forward rates* is introduced to build trajectories of an observable market variable and mimic the effect of a macroeconomic triggering event based on interest rates on the lapse intensity. Contrary to previous works, our *shot-noise intensity* is not constant and the resulting intensity process is not Markovian. Closed-form expressions and analytic sensitivities for the moments of the lapse intensity are provided, showing how lapses can be affected by massive copycat behaviors. Further analyses are then conducted to illustrate how the mean risk varies depending on the model's parameters, while a simulation study compares our results with those obtained using standard practices. The numerical outputs highlight a potential misestimation of the expected number of lapses under extreme scenarios when using classical stress testing methodologies.

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

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).

[†]The work is supported by both the BNP Paribas Cardif Chair "Data Analytics & Models for Insurance" and the ANR Research Project ANR-08-BLAN-0314-01. The views expressed herein are the authors' owns and do not reflect those endorsed by BNP Paribas. Contacts: xavier.milhaud@univ-lyon1.fr, yahia.salhi@univ-lyon1.fr.

1 Introduction

This paper introduces a new mathematical framework in which the modeling of lapse decisions allows to embed relevant effects not properly captured in the existing literature for modeling lapse risk. Naturally arising in real life, these effects are identified as *correlation and contagion among policyholders' behaviors*, and their *dynamic aspect* related to macroeconomic conditions.

Since the 1980s', the lapse risk has become one of the three main risks faced by life insurers, with market and credit risks¹. Policyholders' behaviors have a direct impact on the financial performance of the insurance company and are therefore crucial for the stakeholders of the firm. As stated in [40], most activities of the insurance company are affected by policyholders' behaviors: product design, pricing, reserving, capital allocation, ALM and risk management. We study here the impact of policyholders' behaviors on the solvency of the insurance company: the introduction of fair value accounting, principle-based reserving constraints and *more restrictive regulatory capital requirements* (Solvency II, ORSA) have raised the need for insurance companies to regularly monitor and update policyholders' behavioral assumptions embedded in internal models. Identifying and quantifying the risk factors underlying their insurance products is also of paramount importance (see [40]). To protect the policyholders, insurance regulators require the companies to improve their lapse risk management by performing stress tests (see [6]). The same applies for banks, due to the increasing attention paid by the European Banking Authority [14] on the prepayment risk arising from non-trading activities. In general, appropriate integrated stress testing program should cover scenarios with i) sudden interest rates movements and ii) the breakdown of *behavioral* assumptions. The lapse risk modeling is therefore crucial: regulators need to understand it in order to define adequate capital requirements so as to guarantee the overall stability of the system, and risk managers need to monitor lapse behaviors to prevent large losses that could affect their own activity.

Although this paper assumes a *risk management perspective* and focuses on estimating risk measures associated to losses, it is important to note that lapses can also have a positive impact on the insurer's balance sheet, depending on the economic context and the portfolio under study. For instance, life insurance contracts (with guaranteed return) underwritten more than fifteen years ago are very costly in terms of insurers' reserves, and corresponding lapses are 'good news' for the insurer. From a purely financial viewpoint, the key aspects are the pricing and hedging of the surrender option embedded in most of life insurance contracts, since this option can be exercised at any time.

Nowadays, policyholders and financial markets are becoming increasingly connected, and information about new products are spread out much more quickly. The critical aspect of this fast-evolving area is thus the *dynamic nature of policyholders' behavior*, which results in relationships which are not simple aggregations of underlying stable processes. As a consequence, the modeling of such events is particularly challenging, and dynamic assumptions must be flexible enough to be adapted to the policy's characteristics (specific product, e.g. unit-linked contracts) or to external risk factors.

The question about *How to model lapse behaviors?* has been addressed in various papers, see for instance [34], [22], [21], [20], [28]. Statistical methodologies include classification trees, generalized linear models (GLM), and survival analysis. These techniques enable to integrate risk factors

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

as covariates (e.g. tax relief, contract features, policyholder’s characteristics, firm’s reputation, competition, regulation, and financial markets), thus leading to heterogeneous predictions depending on the policyholder’s profile. However, they globally fail at giving accurate lapse predictions in changing economic environments since policyholders’ behaviors are considered independent, see [27].

In life insurance, the contracts are likely to remain in force for years or decades, which explains why practitioners are used to modeling lapses by distinguishing between *structural* (baseline risk) and *temporary lapses* (related to some disturbance). In the same spirit, [32] has shown through empirical studies that the causes of lapse are globally of two sorts: liquidity needs for personal projects and agents’ rationality, respectively associated to structural and temporary lapses. Policyholders’ personal plans (e.g. purchase a new car) are very difficult to anticipate for the insurer, but they remain quite stable and independent events over time in a large portfolio. On the other hand, agents’ rationality can make policyholders’ decisions become highly correlated in adverse scenarios (e.g. following some experts’ recommendations or largely broadcasted rumors). The main assumption underlying every classical statistical model (independence among individuals) thus becomes violated, causing standard practices (such as GLM) for pricing and reserving to be unadapted. This phenomenon of copycat behaviors and contagion has been observed a few times in bank and insurance sectors, and represents the major threat for insurance companies: the riskiness increases because the policyholders *simultaneously take the same decision before the contractual maturity*. Whatever the situation, i.e. massive lapses or no lapses, the main issue concerns liquidity: either policyholders immediately retrieve their surrender values when lapsing their contract, or the insurer has to pay for guaranteed returns that may lead him to bankruptcy (except if the profitability of the company is higher than these returns, which would never be the case in such a context). To take this into account, regulators have historically defined simple risk management rules to calculate the Solvency Capital Requirement (SCR). Computing the SCR generally requires to assess a baseline risk (structural lapses) to which arbitrary shocks (that stand for temporary lapses) are applied, see for instance the technical specifications of the 5th QIS in the European directive Solvency II². In these recommendations to evaluate the SCR related to the lapse risk in life insurance, the upper-shock of the standard formula consists in adding temporary lapses to the structural part, where the temporary lapse rate can represent up to 30% of the insurance policies with a positive surrender strain. Concerning the development of (partial) internal models, many companies are re-examining their entire actuarial assumption-setting in order to ensure that their risk assessment is appropriate for the environment in which they are currently operating. Most of actuaries currently use a standard deterministic model based on a *S-shaped* curve³ that links the *temporary* lapses to the policyholder’s satisfaction (through the contract return). [25] suggests a stochastic extension to this framework and models the insureds’ 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 artificially integrate the correlation effect, these approaches have some limits since they are ‘static’: shocks are determined *a priori*, and the potential contagion among policyholders’ behaviors in extreme situations is not considered.

²More precisely concerning the lapse risk, see http://archive.eiopa.europa.eu/fileadmin/tx_dam/files/consultations/QIS/QIS5/QIS5-technical_specifications_20100706.pdf, pp.155–159

³See the ONC document by the french regulator (ACPR): https://acpr.banque-france.fr/fileadmin/user_upload/acpr/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).

Here, *correlation and contagion* among policyholders are embedded in the modeling of the lapse intensity process. The novelty of our approach is to introduce *the interest rate dynamics* as the external factor affecting the lapse decision, thus potentially generating temporary surrenders. Our mathematical framework is suitable to model the structural and temporary lapses at the same time under both stressed and unstressed scenarios, still keeping a full analytical tractability of some relevant risk measures. We leverage on a specific extension of Hawkes processes ([17]), the so-called *dynamic contagion process* ([3], [11]). Hawkes processes are a powerful tool applied both in finance ([16]) and insurance ([8, 9, 10]), where the intensity process is piecewise deterministic and enables to integrate the phenomenon of contagion between events ([12]). In our setting, the external shocks (represented by the *shot-noise intensity*) added to the Hawkes-based intensity capture the correlation among policyholders' behaviors when facing adverse economic scenarios. An *external market driven* jump component drives the lapse intensity process depending on the spread between the *market forward interest rates* and the *contractual crediting rate*: when this spread crosses a given threshold, the likelihood of surrenders increases. The risk management perspective assumed in this paper does not aim at considering pricing/hedging issues related to interest rate derivatives, but rather to have a mathematical setting (a *log-normal dynamic* for the *forward rates*) to build trajectories of an observable market variable and mimic the effect of a macroeconomic triggering event based on interest rates on the lapse intensity. Contrary to previous works, our shot-noise intensity is not constant and is derived from an inverse Gaussian distribution: the resulting intensity process is then not Markovian.

The paper is organized as follows: Section 2 discusses the financial setting and describes the proposed mathematical framework by defining the lapse intensity process. Section 3 derives the main theoretical results and provides closed-form expressions for the moments of the lapse intensity; a sensitivity analysis of the long-run mean lapse intensity to each model's parameter is provided and discussed by leveraging on its decomposition in three terms, i.e. *structural*, *market-driven* and *contagion components*. This result is particularly relevant for stress testing and risk monitoring, for example to perform "what-if analysis" under specific scenarios. Further analyses are then performed in Section 4 so as to determine the qualitative impact of the model parameters on the mean risk level for different time horizons. Finally, Section 5 presents a real-world application with numerical results on the number of lapses, associated risk measures and capital requirements: we compare and discuss our results to standard practices (e.g. Solvency II, market practices), and show that huge differences can exist between all aforementioned approaches.

2 Policyholders' behaviors and lapse risk dynamics

As previously mentioned, the lapsation can be triggered and driven either by structural or temporary risk factors: structural risk factors typically 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 to model the policyholders' propensity to lapse by means of a *dynamic contagion process* (see [11]), in order to build an integrated setting allowing to capture both structural and temporary drivers. Besides the influence of structural risk factors, the model focuses on an *interest rate hypothesis* by assuming that the only temporary risk factor

driving lapses is the *spread* between the *market observable interest rate* level and the *contractual crediting rate* defined by the insurance contract. As discussed in [40], lapses are recognized to be negatively correlated to internal rates of returns (such as minimum guaranteed contractual rates) and positively correlated to external rate of returns, such as market interest rates or stock returns (see [7], [22]). From an economic point of view, the reference rate underlying the propensity of policyholders to surrender is identified in the *market observable forward rate*, thus considering the influence of the expectations about future interest rates levels on the lapse intensity dynamics.

Here we consider the scenario in which an investor, before taking his decision to surrender, looks at the forward rates in the market and decides whether to surrender or not depending on the observed return. Of course, the sensitivity of policyholders to interest rate movements differs from one insured to another depending on the contract and market conditions, and big investors (having a huge sum insured) are usually more rational and tend to surrender more often and more quickly. Given that the lapse risk is always material for large contracts, if the insurer is able to anticipate the behavior of such policyholders via the lapse risk modeling, he can prevent huge losses coming from them and their reaction to the forward rate dynamics.

The financial literature for modeling forward rates provides many examples of log-normal dynamics for both the modeling of forward LIBOR rates and forward swap rates ([29], [2], [18] [31], [35], [30], [4]). As stated in [4], one of the most popular families of interest rate models are the market models: the log-normal forward-LIBOR model and the log-normal forward-swap model are the market standards to price cap/floor and swaptions based on Black's formula, i.e. the market quotes liquid cap/floor and swaptions implied volatilities implicitly assuming forward LIBOR and swap rates following a log-normal dynamics.

The introduction of the interest rate dynamics in this paper does not have the aim to price bonds, but rather to build trajectories of an observable market variable in order to mimic and study the impact of a macroeconomic triggering event - based on interest rates - on the lapse intensity. Despite the specific features of each aforementioned model, a log-normal dynamics for the market forward rates is thus introduced into the modeling as a fair representation, at least as first general approximation, of a forward rate dynamics.

2.1 The mathematical model for the lapse intensity

We model the situation in which policyholders decide whether to lapse or not at random (discrete) time intervals, and the probability of a lapsation decision is captured by the hazard function λ_t : the probability of experiencing a lapse in a time interval of length dt is approximately $\lambda_t dt$. The current Subsection is dedicated to the definition and discussion of the stochastic intensity λ_t .

Let us consider the portfolio of policyholders of an insurance company and 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 random times $\{T_i\}_{i=1,2,\dots}$. Now, define the *dynamic contagion process* for the intensity λ_t associated to the counting process N_t as:

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

where

- $\lambda_c \geq 0$ is the constant reversion level,
- $\lambda_0 > 0$ represents the initial constant value (starting point) of the stochastic intensity process,
- $\beta > 0$ is the constant rate of exponential decay,
- $\{X_i\}_{i=1,2,\dots}$ are i.i.d. exponential random variables $X_i \sim \text{Exp}(\gamma), \gamma > 0$, for the magnitude of *self-excited jumps* arriving at random times T_i ,
- $\{Y_j\}_{j=1,2,\dots}$ are i.i.d. exponential random variables $Y_j \sim \text{Exp}(\delta), \delta > 0$, for the magnitude of *external jumps* arriving at random times $\hat{T}_j$, following a counting process $\hat{N}_t$ related to the *market interest rate dynamics*. We assume that the random times $\hat{T}_j$ are hitting times linked to the event in which the *relative gap* RG_t^j between the market interest rates and the insurance company's crediting rate crosses a threshold $B > 0$:

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

with $\hat{T}_0 = 0$, the *generic relative gap* defined as

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

where F_t is the *market observable forward rate* at time t and $R_{\hat{T}_j}^c$ the *contractual crediting rate* at time $\hat{T}_j$. The relative gaps evolve in time with a stochastic dynamics driven by the dynamics of market forward rates, whose evolution is assumed to follow a geometric Brownian motion (GBM) as:

$$\frac{dF_t}{F_t} := \mu dt + \sigma dW_t, \quad F_0 > 0, \quad (4)$$

with $\mu, \sigma \in \mathbb{R}^+$ and $(W_t)_{t \geq 0}$ a standard Brownian motion. Coherently with these assumptions, we derive in Section 3 the model inherent distribution for the inter-arrival times $\Delta \hat{T}_j = \hat{T}_j - \hat{T}_{j-1}$ in order to evaluate the moments of the lapse intensity.

The model in Equation (1) corresponds to a dynamic contagion process (see [11]), i.e. an extension of the classical Hawkes intensity. By using the same notation, the classical Hawkes intensity can be written as

$$\lambda_t = \lambda_c + (\lambda_0 - \lambda_c)e^{-\beta t} + \sum_{i \geq 1} X_i e^{-\beta(t-T_i)} \mathbf{1}_{\{T_i \leq t\}}, \quad (5)$$

corresponding to the first three terms of Equation (1).

Differently from a deterministic Cox-type constant intensity which has been widely used in the actuarial literature for lapses (see [5]), Hawkes-based intensity is stochastic and increases whenever the point process N_t jumps. These *self-excited jumps* of magnitude X_i allow to capture massive lapses generated by the breakdown of classical policyholders' behavioral assumptions. A widespread

panic situation in the market could make policyholders much more inclined to lapse due to correlation and contagion effects (copycat behaviors), and thus making lapse risk much more difficult to monitor for insurance companies, as discussed in [26].

When looking at Equation (5), the policyholder lapsation at time T_i has an impact of magnitude X_i on the intensity process λ_t : this impact exponentially vanishes over time at the (constant) rate β until the mean-reversion level λ_c is reached. Nevertheless, the *temporary risk factor* that can originally cause the *contagion* is *not* explicitly incorporated, while this is the case in Equation (1) thanks to the *external jumps* at random times $\hat{T}_j$.

The dynamic contagion process defined in Equation (1) represents a more general mathematical setting embedding both sources of risk, *self-excited* and *external jumps*, and linking one underlying component to the gap between the *external* market interest rates and the insurance company's crediting rate. From this perspective, the intensity process defined in Equation (1) provides a suitable framework to integrate and capture in the modeling both contagion and correlation effects observed among policyholders' behaviors, linking them to an *observable market component* that can originate and drive contagion. Without loss of generality, we assume that the magnitude of both *self-excited* and *external* jumps is exponentially-distributed with, respectively, parameters γ and δ . The lapse intensity process λ_t therefore depends on both:

- (i) the point process $(N_t)_{t \geq 0}$, introduced as a technical tool to model contagion inside the portfolio, and driven by the *self-excited jumps* at random times T_i (it increases by 1 each time a lapse is registered);
- (ii) an additional *external* point process $(\hat{N}_t)_{t \geq 0}$ related to the *external jumps* arriving at random times $\hat{T}_j$: $\hat{N}_t = \sum_{j \geq 1} \mathbf{1}_{\{\hat{T}_j \leq t\}}$ and linked to the *dynamics of market observable forward interest rates*.

Remark 1 Notice that the stochastic intensity process λ_t defined in Equation (1) is always above the constant reversion level λ_c , i.e. $\lambda_t \in [\lambda_c, \infty)$. The following special cases are discussed:

- if $\lambda_c = 0$, the distribution of λ_t converges to the distribution of a degenerate random variable at 0 as $t \rightarrow \infty$;
- if $\lambda_0 = \lambda_c$, the term $(\lambda_0 - \lambda_c) \cdot e^{-\beta t}$ disappears, representing the scenario in which the starting point of the system already equals the constant reversion level, i.e. no further mean reversion adjustments are needed. In this case, the long run level of the intensity driven by the structural risk factors component has already been reached.

In order to describe the component linked to the *market observable forward rates*, 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. When policyholders have to decide whether to keep the insurance contract or lapse it, they are often inclined to compare it with other insurance or financial products: most policyholders base their lapsation decision on how attractive they deem their policy w.r.t. some alternatives in the market, rather than on a pure 'fair value' logic underlying their option. Thus, the insurer must pay attention to the way the products are presented and to the available alternatives on the market.

All along the contract lifetime, policyholders usually compare their *contractual crediting rate* $((R_t^c)_{t \geq 0})$ to some benchmark mid/long-term interest rate observed in the market, namely the *market forward rate* $(F_t)_{t \geq 0}$ as proxy of a representative benchmark rate.

The *contractual crediting 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 *contractual crediting rate* and the *market observable interest rate levels*. How can the market interest rate trajectory F_t impact the lapse risk? A rational policyholder is obviously more inclined to lapse as the *spread* between the market interest rate and the crediting rate increases. Let us indicate this spread at time $t = 0$ as the first relative gap RG_t^0 , defined as:

$$RG_t^0 := \frac{F_t - R_0^c}{R_0^c}, \quad (6)$$

where F_t is the *market forward interest rate* at time t , and $R_0^c := R^g$ coincides with the guaranteed return 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, $RG_0^0 \approx 0$ for market clearing competitive reasons, then F_t at least equals R^g since the life insurance contract should offer more guarantees than the benchmark market rate. The same relative gap can be defined for each point in time, and the company can adjust the contractual crediting 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*. This should be particularly relevant for “big” investors, i.e. policyholders having a huge sum insured. We can expect big investors being more responsive to changes in the market observable forward rates than “small” customers, and their lapsation more relevant for the insurance company from both a liquidity and financial perspective. Let us assume that subsequent adjustments of the contractual crediting 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. $R_{\hat{T}_1}^c = F_{\hat{T}_1}$. As a consequence, the new relative gap RG_t^1 is given by

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

If the insurer follows this mechanism to set the *contractual crediting 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 of hitting random times $(\hat{T}_j)_{j=0,1,\dots}$ satisfying

$$\hat{T}_{j+1} := \inf\{t > \hat{T}_j, RG_t^j = B\},$$

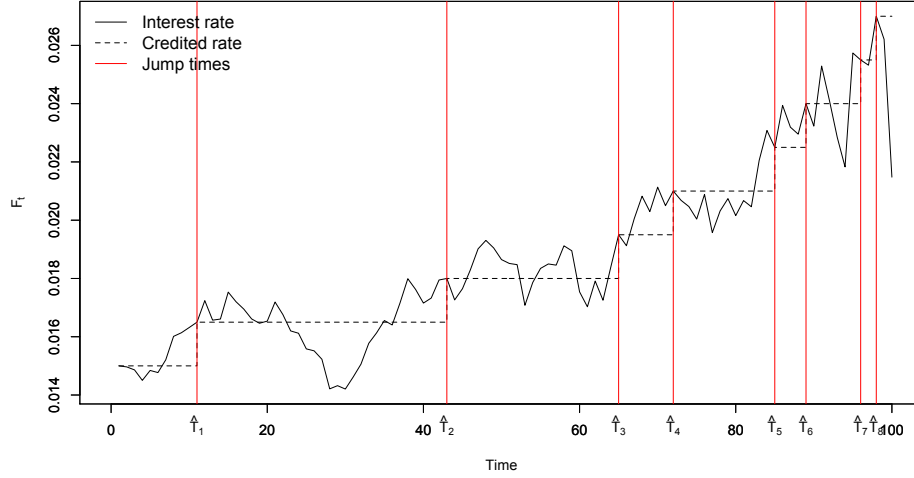


Figure 1: **Market interest rate dynamics and contract credited rate adjustments.** Simulated path of a GBM with $F_0 = 1.5\%$, $\mu = 5\%$, $\sigma = 30\%$. The random times $\hat{T}_j$ at which the contractual crediting rate R_t^c is adjusted are reported for an exogenous threshold set to $B = 10\%$.

and associated to the generic relative gap

$$RG_t^j = \frac{F_t - R_{\hat{T}_j}^c}{R_{\hat{T}_j}^c} = \frac{F_t - F_{\hat{T}_j}}{F_{\hat{T}_j}}, \quad \hat{T}_j \leq t < \infty. \quad (7)$$

Figure 1 illustrates the updating mechanism of the *contractual crediting rate* for a simulated path of the market forward interest rate F_t , whose dynamics follows a GBM defined in Equation (4). The adjustments of the contractual crediting rate are not instantaneous in practice, thus we deem reasonable and more realistic to consider that the propensity of policyholders to lapse can jump (see the last term in Equation (1)). In particular, we suppose it to jump at random times $\hat{T}_j$, and then to vanish over time thanks to subsequent efforts made by the insurer.

Figure 2 shows a typical trajectory of the proposed dynamic contagion process for the intensity λ_t defined in Equation (1): one trajectory of the underlying market interest rate dynamics driving the external jumps is reported in Figure 1.

Remark 2 Notice that N_t does not (necessarily) jump at random times $\hat{T}_j$, since we are not supposing to have a lapse each time the threshold B is crossed by the relative gap between the interest rate and the contractual credited rate. Equation (1) is rather modeling a more general case: each time the spread between market interest rates and the contractual crediting rate crosses the threshold, a jump in the intensity is registered, thus injecting volatility via additional internal turbulences in the dynamic contagion process (upward movement driven by the external jump component). In such a case, the counting process $\hat{N}$ increases by 1. The counting process N_t instead jumps only in case of lapses: it can jump at time $\hat{T}_j$ only if a policy termination is registered at the same time, thus in the case of $\hat{T}_j = T_k$ for some (j, k) . If no policy termination is registered at time $\hat{T}_j$, the unique effect of $RG_t^j > B$ will be an external jump in the intensity.

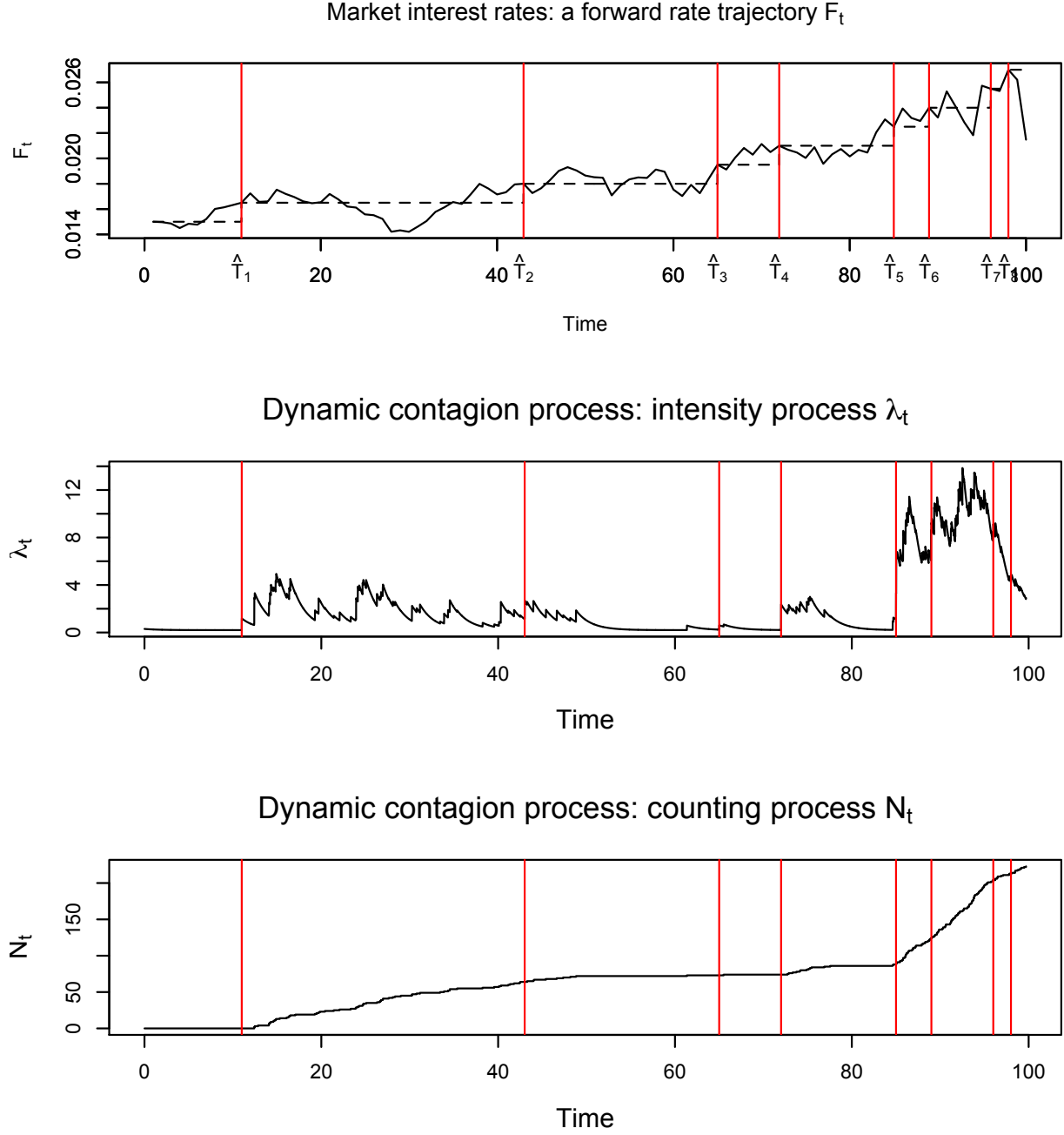


Figure 2: **Intensity dynamics λ_t and counting process N_t .** The two first graphs respectively show one particular sample path of the interest rate and the corresponding intensity λ_t given in Equation (1). Vertical lines stand for the *external* jump times $\hat{T}_j$. The third graph is associated 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 $\lambda_0 = 0.5, \lambda_c = 0.4, \beta = 2, \delta = 2, \gamma = 1.5$.

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 λ_t defined in Equation (1). 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 random times $\hat{T}_j$ associated to jumps in the intensity.

3.1 Market interest rate dynamics and external jumps

The random times $(\hat{T}_j)_{j=0,1,\dots}$ are hitting times of the process describing the evolution of the relative gaps RG_t^j defined in Equation (7). Our methodology is developed under the assumption of a specific stochastic dynamics for the interest rate. Here, a GBM with drift μ and diffusion coefficient σ is considered for the market forward rates (see Equation (4)), whose solution is given by

$$F_t = F_0 \cdot e^{(\mu - \sigma^2/2)t + \sigma W_t}. \quad (8)$$

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:

$$\hat{T}_j = \hat{T}_{j-1} + \inf\{t > 0, (\mu - \sigma^2/2)t + \sigma W_t = \log(1 + B)\}, \quad (9)$$

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 (see [39]). 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

$$g(t) = \left(\frac{\theta_2}{2\pi t^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{\theta_2(t - \theta_1)^2}{2\theta_1^2 t}\right\}, \quad (10)$$

$$G(t) = \Phi\left(\sqrt{\frac{\theta_2}{t}}\left(\frac{t}{\theta_1} - 1\right)\right) + \exp\left\{\frac{2\theta_2}{\theta_1}\right\} \Phi\left(-\sqrt{\frac{\theta_2}{t}}\left(\frac{t}{\theta_1} + 1\right)\right), \quad (11)$$

with $\Phi(\cdot)$ being the standard normal CDF (see [38, 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 $\mathbb{P}(\hat{T}_j \leq t) = G^{j*}(t)$ where G^{j*} is the j -fold convolution of G . Using $\mathbb{E}[\hat{N}_t] = \sum_{j=0}^{\infty} \mathbb{P}(\hat{N}_t \geq j)$, we have

$$h(t) = \mathbb{E}[\hat{N}_t] = \sum_{j=0}^{\infty} G^{j*}(t). \quad (12)$$

As shown in [38], the CDF G^{j*} is still an inverse Gaussian, with mean $j\theta_1$ and shape $j\theta_2$.

3.2 Recursive formulas for moments derivation

We derive here the moments of both the intensity λ_t and 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^{th} derivative of m with respect to θ , such that $m^{(n)}(t, 0)$ is the n^{th} moment of λ_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

$$Z_t = \sum_{i=1}^{N_t} X_i e^{\beta T_i} \quad \text{and} \quad \hat{Z}_t = \sum_{i=1}^{\hat{N}_t} Y_i e^{\beta \hat{T}_i}. \quad (13)$$

These processes are discounted compounded renewal processes (see [24]), and $\xi^{(n)}(t, \theta)$ and $\hat{\xi}^{(n)}(t, \theta)$ refer to the n^{th} derivative of $\xi(t, \theta)$ and $\hat{\xi}(t, \theta)$ with respect to θ . Hence, the process λ_t can be written in the following form

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

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

Proposition 1 *The MGF of λ_t can be decomposed as follows:*

$$m(t, \theta) = e^{\theta(\lambda_c + (\lambda_0 - \lambda_c)e^{-\beta t})} \xi(t, \theta e^{-\beta t}) \hat{\xi}(t, \theta e^{-\beta t}), \quad (14)$$

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 (see [24]).

Lemma 1 *The MGFs ξ , $\hat{\xi}$ 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*

$$\xi(t, \theta) = 1 + \int_0^t \left(\frac{\theta e^{\beta u}}{\gamma - \theta e^{\beta u}} \right) \xi(t - u, \theta e^{\beta u}) m^{(1)}(u, 0) du, \quad (15)$$

$$\hat{\xi}(t, \theta) = 1 + \int_0^t \left(\frac{\theta e^{\beta u}}{\delta - \theta e^{\beta u}} \right) \hat{\xi}(t - u, \theta e^{\beta u}) dh(u), \quad (16)$$

where $h(t)$ is given in (12). The moments of Z_t , $\hat{Z}_t$ have the following form:

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

$$\hat{\xi}^{(n)}(t, 0) = \sum_{k=0}^{n-1} \frac{n!}{k!} \frac{1}{\delta^{n-k}} \int_0^t e^{n\beta u} \hat{\xi}^{(k)}(t - u, 0) dh(u), \quad (18)$$

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 by resorting to techniques based on the infinitesimal generator of a Markov process and martingale arguments, as in [11]. In their paper, the exponential distribution of the inter-arrival times of external jumps makes the process λ_t be Markovian, which is not the case in our setting. Even if the lapse intensity λ_t defined in Equation (1) is not a Markov process, it is possible to transform it into a Markovian one by introducing as supplementary process, *the time elapsed since the last external jump* (or the time remaining to the next one), referring to the method described in [15]. In order to achieve this result, we rely on a differential argument based on the recursive formulas reported in 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^{th} derivative of the lapse intensity MGF is given recursively as follows:*

$$\begin{aligned} m^{(n)}(t, \theta) &= (\lambda_c + (\lambda_0 - \lambda_c)e^{-\beta t})m^{(n-1)}(t, \theta) \\ &\quad + \sum_{i=0}^{n-1} \binom{n-1}{i} e^{-(n-i)\beta t} \left(I_i(t, \theta) + \widehat{I}_i(t, \theta) \right) m^{(i)}(t, \theta), \end{aligned} \quad (19)$$

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 λ_t by letting $\theta = 0$ in Equation (19). 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 λ_t is given by*

$$\mathbb{E}[\lambda_t] := m^{(1)}(t, 0) = \left(\lambda_0 - \frac{\beta \lambda_c}{\beta - 1/\gamma} \right) e^{-(\beta - \frac{1}{\gamma})t} + \frac{\beta \lambda_c}{\beta - 1/\gamma} + \frac{1}{\delta} \int_0^t e^{-(\beta - \frac{1}{\gamma})(t-s)} h'(s) ds, \quad (20)$$

and is the solution of the differential equation

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

with $h(t)$ given in Equation (12).

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 can write:

$$\mathbb{E}[N_t] = \mathbb{E} \left[\int_0^t \lambda_s ds \right] = \int_0^t m^{(1)}(s, 0) ds, \quad (21)$$

where Fubini-Tonelli's theorem has been used.

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

$$\mathbb{E}[\lambda_t] = \left(\lambda_0 - \frac{\beta \lambda_c}{\beta - 1/\gamma} \right) e^{-(\beta - \frac{1}{\gamma})t} + \frac{\beta \lambda_c}{\beta - 1/\gamma} + \frac{1}{\delta} \left(\sum_{j \geq 1} \int_0^t g^{j*}(s) e^{-(\beta - \frac{1}{\gamma})(t-s)} ds \right). \quad (22)$$

The numerical calculation of (22) is closely related to the computation of the infinite sum

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

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 θ_1 and θ_2 . The inverse Gaussian density g^{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 (θ_1, θ_2) are the parameters of the inverse Gaussian law given in (10). 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*

$$\begin{aligned} \mathbb{E}[\lambda_t] = & \left(\lambda_0 - \frac{\beta \lambda_c}{\beta - 1/\gamma} \right) e^{-(\beta - \frac{1}{\gamma})t} + \frac{\beta \lambda_c}{\beta - 1/\gamma} + \frac{1}{\delta} e^{-(\beta - \frac{1}{\gamma})t} \times \\ & \sum_{j \geq 1} e^{j \frac{\theta_2}{\theta_1}} \left[e^{-j\sqrt{k\theta_2}} \Phi \left(\frac{t\sqrt{k} - j\sqrt{\theta_2}}{\sqrt{t}} \right) + e^{j\sqrt{k\theta_2}} \Phi \left(-\frac{t\sqrt{k} + j\sqrt{\theta_2}}{\sqrt{t}} \right) \right], \end{aligned} \quad (23)$$

(ii) *else*

$$\mathbb{E}[\lambda_t] = \left(\lambda_0 - \frac{\beta \lambda_c}{\beta - 1/\gamma} \right) e^{-(\beta - \frac{1}{\gamma})t} + \frac{\beta \lambda_c}{\beta - 1/\gamma} + \frac{1}{\delta} \sum_{j \geq 1} e^{-\frac{\theta_2}{\theta_1^2} \frac{(j\theta_1 - t)^2}{2t}} \times \text{Re}(w(z_j)), \quad (24)$$

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

Proof: see Appendix A.5. ■

As we can notice from Equations (23)-(24), the mean lapse intensity at a generic time t can be decomposed as the sum of three terms: the first two components are only affected by structural and contagion risk factors, via parameters $\beta, \gamma, \lambda_0, \lambda_c$, while the last term involves all parameters defining the market-driven component linked to the external jumps and thus to the interest rate dynamics and the distribution of inter-arrival times.

We are now interested in deriving the *limiting behavior* (as $t \rightarrow \infty$) 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 [3] for general decay functions, as well as the externally excited framework considered in [11]). 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 λ_t .

Proposition 4 *Under the condition $\beta\gamma > 1$, the limiting expected lapse intensity $\bar{\lambda}_\infty^E$ can be written as follows:*

$$\bar{\lambda}_\infty^E := \lim_{t \rightarrow \infty} \mathbb{E}[\lambda_t] = \frac{\beta\lambda_c}{\beta - 1/\gamma} + \frac{1}{\delta\theta_1(\beta - 1/\gamma)}, \quad (25)$$

with $\mathbb{E}[\lambda_t]$ given in Equation (22).

Proof: see Appendix A.6. ■

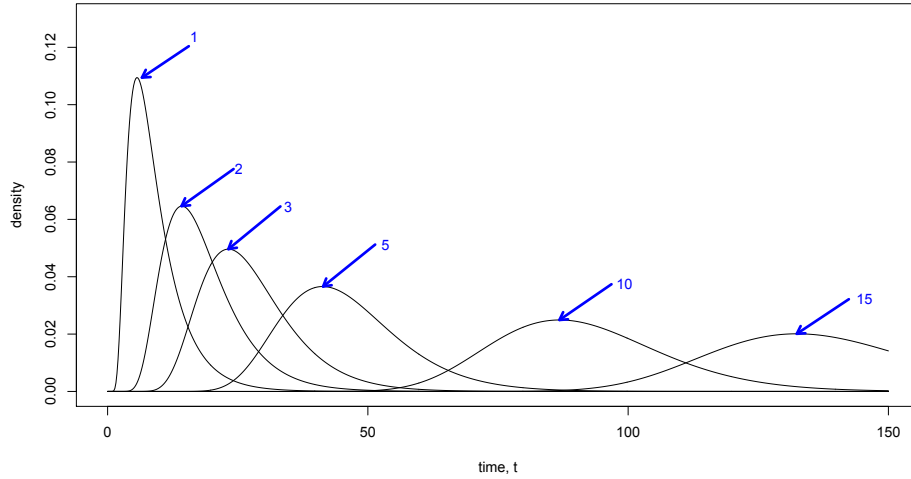


Figure 3: **Density $g^{j*}(t)$ of the sum of j i.i.d. inverse Gaussian random variables.** The graph reports the behavior of the density function $g^{j*}(t)$ of an inverse Gaussian law for different values of $j \in \{1, 2, 3, 5, 10, 15\}$, where the density function $g(t)$ is given in Equation (10).

Let us discuss the result given in Equation (25): the expectation of the lapse intensity as time $t \rightarrow \infty$ can be disentangled in distinct components as:

$$\bar{\lambda}_\infty^E := \frac{\bar{\lambda}_\infty^S + \bar{\lambda}_\infty^M}{\bar{\lambda}_\infty^C}, \quad (26)$$

where

$$\bar{\lambda}_\infty^S := \beta \lambda_c, \quad (27)$$

$$\bar{\lambda}_\infty^M := \frac{1}{\delta \theta_1} = \frac{\mu - \frac{1}{2}\sigma^2}{\delta \log(1+B)}, \quad (28)$$

$$\bar{\lambda}_\infty^C := (\beta - 1/\gamma), \quad (29)$$

represent, respectively, the *structural component*, the *market-driven component* and the *contagion component* (in Equation (28) we have used $\theta_1 := 2 \log(1+B)/(2\mu - \sigma^2)$).

Let us consider the *structural component* $\bar{\lambda}_\infty^S$: it is uniquely related to structural risk factors and self-excited jumps driving the lapse dynamics. This term is function of the constant reversion level λ_c and the rate of exponential decay β . Besides the effects of the structural risk factors, an adjustment to the lapse intensity is assumed to come from the *market-driven component* $\bar{\lambda}_\infty^M$, since interest rate levels are driving the intensity via the external jumps by injecting additional internal turbulence in the modeling. As we can see from Equation (28), this market-driven component is an upward correction (under $\mu - \frac{1}{2}\sigma^2 > 0$) for the mean intensity, which is function of the size of external jumps δ , the drift μ and the diffusion coefficient σ of the market forward interest rate dynamics, the threshold B capturing the movements of the relative gap between market interest rates and the contractual crediting rate. Observe that $\text{sign}(\mu - \frac{1}{2}\sigma^2)$ is affecting the sign of $\bar{\lambda}_\infty^M$, by representing the combination of drift and convexity adjustment for the forward rate dynamics. In case of $\mu - \frac{1}{2}\sigma^2 > 0$, forward rates are increasing, thus the relative gap is more likely to cross the threshold B by generating internal turbulence into the lapse intensity via the external jumps. In case of $\mu - \frac{1}{2}\sigma^2 < 0$, we must have:

$$\mu - \frac{1}{2}\sigma^2 \geq -\beta \lambda_c \cdot \delta \log(1+B) \quad (30)$$

coming from $\bar{\lambda}_\infty^S + \bar{\lambda}_\infty^M \geq 0$, in order to guarantee the positiveness of the expected mean lapse intensity $\bar{\lambda}_\infty^E$. The *contagion component* $\bar{\lambda}_\infty^C$ is instead driven by parameters β, γ , thus having the rate of exponential decay and the mean size of self-excited jumps as underlying risk factors. The positiveness of the contagion component is guaranteed by the *stationary condition* $\beta\gamma > 1$. Tables 1 and 2 summarize the effect of each model's parameter on the mean lapse intensity in the long run and disentangle the sensitivity by highlighting the re-activeness of the *structural*, *market-driven* and *contagion components*.

Remark 3 Observe that the starting point of the intensity process λ_0 does not enter neither in the long-run structural component $\bar{\lambda}_\infty^C$ nor in the long-run market-driven component $\bar{\lambda}_\infty^M$, while it is the case for short and mid time horizons, see Equations (23)-(24). When $t < \infty$, subsequent adjustments to the long run mean must be done, thus fluctuations around the long run mean depends on the starting point; when considering $t \rightarrow \infty$ we are implicitly studying a long run equilibrium convergence level, that is why the starting point plays no role in $\bar{\lambda}_\infty^E$. Notice that if $\lambda_c = 0$ we have a null structural component $\bar{\lambda}_\infty^S = 0$, meaning that the expected lapse intensity only depends on the market-driven and contagion components, i.e. Equation (26) simplifies to $\bar{\lambda}_\infty^E := \bar{\lambda}_\infty^M / \bar{\lambda}_\infty^C$.

4 Lapse intensity: sensitivity to the model parameters

The aim of this Section is to study the role played by each model's parameter in the proposed mathematical framework by highlighting its influence on the lapse risk. In order to analyze and measure this impact from both a qualitative and quantitative perspectives, we consider the expectation of the intensity process $\mathbb{E}[\lambda_t]$ as the main risk indicator, see Proposition 3 and Figure 4 (base case). We focus hereafter on the expectation $\mathbb{E}[\lambda_t]$ for three main reasons: i) 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; ii) it enables to derive other qualitative and quantitative results (such as the average number of lapses) quite easily; iii) its sensitivity to each model's parameter can be derived in analytic closed form. This is what we do by considering the long-run mean intensity given in Equation (25) for the expected mean lapse intensity as $t \rightarrow \infty$. In particular, *sensitivity* and *elasticity* to the model's parameters are derived for each component in which the lapse intensity can be disentangled in, i.e. the *structural component* $\bar{\lambda}_\infty^S$, the *market-driven component* $\bar{\lambda}_\infty^M$, the *contagion component* $\bar{\lambda}_\infty^C$.

Let us have a look to Figures 4 and 5 for a visual insight of the qualitative behaviour of $\mathbb{E}[\lambda_t]$. Table 4 summarizes the directional impacts observed in the mean lapse intensity for a change in the model's parameters by considering different time horizons while Tables 1-3 focus on the long term mean lapse intensity, i.e. $\bar{\lambda}_\infty^E$ given in Equation (25) and its components $\bar{\lambda}_\infty^S, \bar{\lambda}_\infty^M, \bar{\lambda}_\infty^C$ (structural, market driven, contagion) by reporting the analytic expression of the sensitivity to the model's parameters and the corresponding elasticity.

	Structural component	Market-driven component	Contagion component	Long-term expected intensity
Model parameter	$\frac{\partial \bar{\lambda}_\infty^S}{\partial p}$	$\frac{\partial \bar{\lambda}_\infty^M}{\partial p}$	$\frac{\partial \bar{\lambda}_\infty^C}{\partial p}$	$\frac{\partial \bar{\lambda}_\infty^E}{\partial p}$
p				
λ_0				
λ_c	$\frac{\partial \bar{\lambda}_\infty^S}{\partial \lambda_c} = \beta > 0$			$\frac{\partial \bar{\lambda}_\infty^E}{\partial \lambda_c} = \frac{\partial \bar{\lambda}_\infty^S}{\partial \lambda_c} \frac{1}{\bar{\lambda}_\infty^C}$
β	$\frac{\partial \bar{\lambda}_\infty^S}{\partial \beta} = \lambda_c \geq 0$		$\frac{\partial \bar{\lambda}_\infty^C}{\partial \beta} = 1 > 0$	$\frac{\partial \bar{\lambda}_\infty^E}{\partial \beta} = \frac{1}{\bar{\lambda}_\infty^C} \left(\frac{\partial \bar{\lambda}_\infty^S}{\partial \beta} \frac{1}{\bar{\lambda}_\infty^C} - \bar{\lambda}_\infty^E \right)$
γ			$\frac{\partial \bar{\lambda}_\infty^C}{\partial \gamma} = \frac{1}{\gamma^2} > 0$	$\frac{\partial \bar{\lambda}_\infty^E}{\partial \gamma} = -\frac{\partial \bar{\lambda}_\infty^C}{\partial \gamma} \cdot \frac{\bar{\lambda}_\infty^S + \bar{\lambda}_\infty^M}{(\bar{\lambda}_\infty^C)^2}$
μ		$\frac{\partial \bar{\lambda}_\infty^M}{\partial \mu} = \frac{1}{\delta \log(1+B)} > 0$		$\frac{\partial \bar{\lambda}_\infty^E}{\partial \mu} = \frac{\partial \bar{\lambda}_\infty^M}{\partial \mu} \frac{1}{\bar{\lambda}_\infty^C}$
σ		$\frac{\partial \bar{\lambda}_\infty^M}{\partial \sigma} = -\frac{\sigma}{\delta \log(1+B)} < 0$		$\frac{\partial \bar{\lambda}_\infty^E}{\partial \sigma} = \frac{\partial \bar{\lambda}_\infty^M}{\partial \sigma} \frac{1}{\bar{\lambda}_\infty^C}$
B		$\frac{\partial \bar{\lambda}_\infty^M}{\partial B} = -\frac{\mu - \frac{1}{2}\sigma^2}{\delta(1+B)\log^2(1+B)}$		$\frac{\partial \bar{\lambda}_\infty^E}{\partial B} = \frac{\partial \bar{\lambda}_\infty^M}{\partial B} \frac{1}{\bar{\lambda}_\infty^C}$
δ		$\frac{\partial \bar{\lambda}_\infty^M}{\partial \delta} = -\frac{1}{\delta^2} \frac{(\mu - \frac{1}{2}\sigma^2)}{\log(1+B)}$		$\frac{\partial \bar{\lambda}_\infty^E}{\partial \delta} = \frac{\partial \bar{\lambda}_\infty^M}{\partial \delta} \frac{1}{\bar{\lambda}_\infty^C}$

Table 1: **Long term horizon: sensitivity of the expected lapse intensity on the model's parameters.** Analytic expressions of the sensitivity of the expected intensity on each models' parameter p . The effect on $\bar{\lambda}_\infty^E$ is reported as function of the structural, market-driven and contagion components, namely $\bar{\lambda}_\infty^S, \bar{\lambda}_\infty^M, \bar{\lambda}_\infty^C$. The sensitivity of the structural component, the market-driven component and the contagion component in which the intensity can be disentangled are reported.

	Long term mean intensity level $\bar{\lambda}_\infty^E := \mathbb{E}[\lambda_t]$ in Equation (25).	
Model parameter	$\frac{\partial \bar{\lambda}_\infty^E}{\partial p}$	$sign(\frac{\partial \bar{\lambda}_\infty^E}{\partial p})$
p		
λ_c	$\frac{\partial \bar{\lambda}_\infty^E}{\partial \lambda_c} = \frac{\beta}{\beta - \frac{1}{\gamma}}$	> 0
β	$\frac{\partial \bar{\lambda}_\infty^E}{\partial \beta} = -\frac{1}{(\beta - \frac{1}{\gamma})^2} \left(\frac{\lambda_c}{\gamma} + \frac{\mu - \frac{1}{2}\sigma^2}{\delta \log(1+B)} \right)$	$-sign(\frac{\lambda_c}{\gamma} + \frac{\mu - \frac{1}{2}\sigma^2}{\delta \log(1+B)})$
γ	$\frac{\partial \bar{\lambda}_\infty^E}{\partial \gamma} = -\frac{\bar{\lambda}_\infty^S + \bar{\lambda}_\infty^M}{\gamma^2(\beta - \frac{1}{\gamma})^2}$	< 0
μ	$\frac{\partial \bar{\lambda}_\infty^E}{\partial \mu} = \frac{1}{\delta(\beta - \frac{1}{\gamma}) \log(1+B)}$	> 0
σ	$\frac{\partial \bar{\lambda}_\infty^E}{\partial \sigma} = -\frac{\sigma}{\delta(\beta - \frac{1}{\gamma}) \log(1+B)}$	< 0
B	$\frac{\partial \bar{\lambda}_\infty^E}{\partial B} = -\frac{(\mu - \frac{1}{2}\sigma^2)}{\delta(\beta - \frac{1}{\gamma})} \frac{1}{(1+B) \log^2(1+B)}$	$sign(\frac{1}{2}\sigma^2 - \mu)$
δ	$\frac{\partial \bar{\lambda}_\infty^E}{\partial \delta} = -\frac{1}{\delta^2} \frac{(\mu - \frac{1}{2}\sigma^2)}{(\beta - \frac{1}{\gamma}) \log(1+B)}$	$sign(\frac{1}{2}\sigma^2 - \mu)$

Table 2: **Long term horizon: sensitivity of the expected lapse intensity on the model's parameters.** The table reports the analytic expression of the sensitivity of the expected lapse intensity on each models' parameter, generally indicated as p together with a study of its sign.

	structural component	market-driven component	contagion component
Model parameter	$\eta(\bar{\lambda}_\infty^S, p) := \frac{\partial \bar{\lambda}_\infty^S}{\partial p} \cdot \frac{p}{\bar{\lambda}_\infty^S}$	$\eta(\bar{\lambda}_\infty^M, p) := \frac{\partial \bar{\lambda}_\infty^M}{\partial p} \cdot \frac{p}{\bar{\lambda}_\infty^M}$	$\eta(\bar{\lambda}_\infty^C, p) := \frac{\partial \bar{\lambda}_\infty^C}{\partial p} \cdot \frac{p}{\bar{\lambda}_\infty^C}$
p			
λ_c	$\eta(\bar{\lambda}_\infty^S, \lambda_c) = 1$		
β	$\eta(\bar{\lambda}_\infty^S, \beta) = 1$		$\eta(\bar{\lambda}_\infty^C, \beta) = \frac{\beta}{\beta - \frac{1}{\gamma}} > 1$
γ			$\eta(\bar{\lambda}_\infty^C, \gamma) = \frac{1}{\gamma(\beta - \frac{1}{\gamma})}$
μ		$\eta(\bar{\lambda}_\infty^M, \mu) = \frac{\mu}{\mu - \frac{1}{2}\sigma^2}$	
σ		$\eta(\bar{\lambda}_\infty^M, \sigma) = -\frac{\sigma^2}{\mu - \frac{1}{2}\sigma^2}$	
B		$\eta(\bar{\lambda}_\infty^M, B) = -\frac{B}{(1+B) \log(1+B)}$	
δ		$\eta(\bar{\lambda}_\infty^M, \delta) = -1$	

Table 3: **Elasticity of the lapse intensity components on the model's parameters.** The Table reports the analytic expression of the elasticity of the lapse intensity components on each models' parameter, generally indicated as p . The elasticity is reported, separately, for the structural, market-driven and contagion components; the elasticity of each component w.r.t. the generic model parameter p is denoted as $\eta(\bar{\lambda}_\infty^S, p)$, $\eta(\bar{\lambda}_\infty^M, p)$, $\eta(\bar{\lambda}_\infty^C, p)$.

Since the main novelty of the proposed mathematical framework is the introduction of a macroeconomic variable into the lapse risk modeling, let us start by considering how the market parameters affect the first moment of the intensity process λ_t . In particular, let us start by analyzing the *external/market-driven component* $\bar{\lambda}_\infty^M$ and analyze how changes of the GBM parameters affect the first moment of the intensity process λ_t . First, look at the drift μ : according to Figure 5, a greater drift corresponds to a higher mean intensity process for each time horizon. A greater drift naturally makes the exogenous barrier B be hit more often (everything else fixed), thus increasing the external risk component by generating the occurrence of additional external jumps. This is because we are assuming the market forward rates growing at a higher rate. 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 μ . As we can see from Tables 1-3, the drift parameter only affects the *market-driven component* $\bar{\lambda}_\infty^M$ which has a positive sensitivity to this parameter and an elasticity bigger than 1 (under $\mu - \frac{1}{2}\sigma^2 > 0$). On the contrary, the volatility σ reveals to have an impact on the time needed by the mean intensity to reach its long-run equilibrium $\bar{\lambda}_\infty^E$: the more σ increases, the longer this time is. Increasing the volatility in the interest rate dynamics means increasing the level of uncertainty characterizing the system, and more uncertainty obviously postpones the stability by injecting turbulence in the system. In mathematical terms, this can be seen by looking at Equation (25) and Proposition

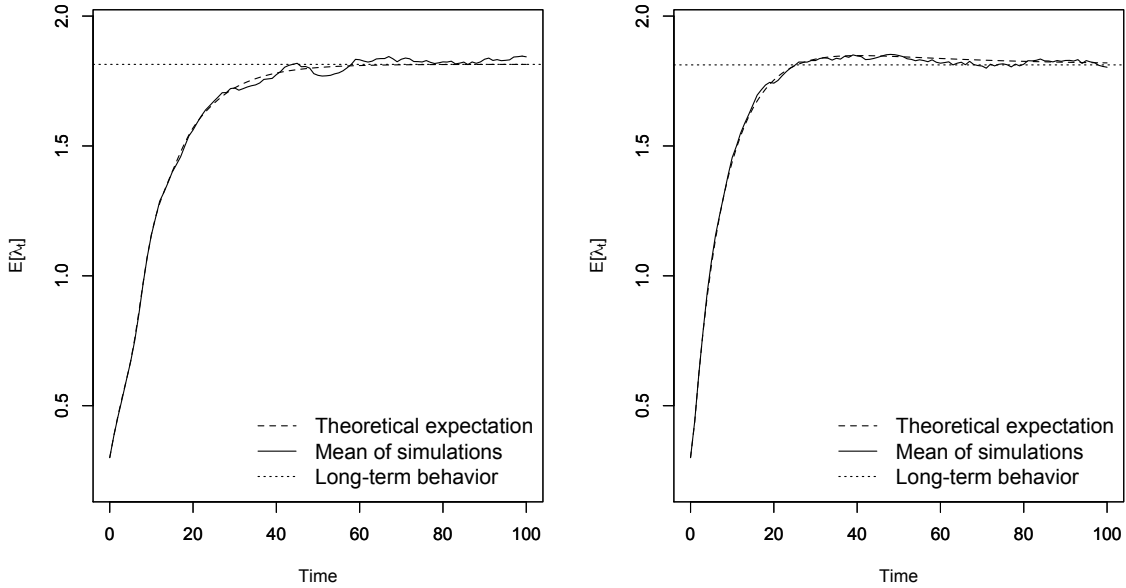


Figure 4: **Expectation of the lapse intensity process for $t \in [0, 100]$.** The plots compare the outcome of a Monte Carlo simulation ($2 \cdot 10^4$ trajectories) to the theoretical expressions of Proposition 3. On the *left* the case of Equation (23), with parameters: $F_0 = 1.5\%$, $\mu = 1\%$, $\sigma = 1\%$, $B = 10\%$, $\lambda_0 = 0.3$, $\lambda_c = 0.2$, $\delta = 1.5$, $\gamma = 2$, and $\beta = 0.6$. The plot on the *right* corresponds to the case of Equation (24), with the same values of parameters except that $\sigma = 5\%$. The two plots also report (dotted line) the long term value of the lapse intensity $\bar{\lambda}_\infty^E$ derived in Equation (25). We have $\bar{\lambda}_\infty^E = 1.89$ on the left and $\bar{\lambda}_\infty^E = 1.81$ on the right.

3: increasing σ makes the mean intensity process raising in the short run and decreasing in the long-term. As for the drift μ , the volatility parameter σ only affects $\bar{\lambda}_\infty^M$: the negative sensitivity of the *market-driven component* w.r.t. σ reported in Table 1 corresponds to a negative sensitivity of the resulting long-run equilibrium level $\bar{\lambda}_\infty^E$ (see Table 2).

Let us now focus on the exogenous barrier B defined for the relative gaps RG_t^j capturing the *spread* between the contractual crediting 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 effect on the mean intensity process is. A higher number of external jumps will instead occur if B decreases, making the intensity greater via an interest rate *market-driven* effect. Tables 1-3 show that the sign of both sensitivity and elasticity w.r.t. this parameter strictly depends on the sign of $\mu - \frac{1}{2}\sigma^2$, i.e. on the growth rate and convexity adjustment underlying the forward rate dynamics. The *external/market-driven component* of the intensity process is also affected by parameter δ . Changing δ produces an impact similar to the one produced by a bump in the exogenous barrier B , since the ratio $1/\delta$ directly relates to the mean size of external jumps (as we can see from the analytic expressions of the sensitivities reported in Tables 1 and 2).

Let us consider now the parameters of the *self-excited component* $\bar{\lambda}_\infty^C$: the exponential decay β has one of the most relevant impacts. This parameter represents the speed at which the intensity λ_t decreases once a jump has occurred. Not surprisingly, a low β 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 time horizon (see Table 4). As time passes, a low β increases the gap with respect to the base case intensity process, highlighting that the effect of β clearly accumulates over time and leads to a much higher long-term mean intensity. As we can see from Tables 1-3, the contagion component has a positive constant derivative w.r.t. β and an elasticity > 1 , meaning that a 1% increase in β translates into a bigger %–change in the component $\bar{\lambda}_\infty^C$. The same argument applies when considering a change in $\bar{\lambda}_\infty^C$ after a variation of γ parameter, which is in line with the stationary condition: when β 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 γ gets higher, the mean size of internal jumps decreases, so that the mean lapse intensity process $\bar{\lambda}_\infty^E$ is lower, as well as the elapsed time before reaching the stationary level. Recall that these two parameters β, γ capture the self-excited component of the proposed dynamic contagion process.

Finally, the mean lapse intensity has a dependence on parameters λ_0, λ_c via the *structural component*. The initial intensity level λ_0 has no long-term effect on the mean intensity level (no sensitivity of $\bar{\lambda}_\infty^E$ to this parameter) and on the time to reach the stationary regime; this parameter only affects the short term behaviour of the intensity process. Besides these effects driven by the starting point of the process, the mean reversion level λ_c obviously impacts not only the short term but also the long-term mean intensity. This is consistent with the analytic expression given in Equation (22), where it is evident that $\mathbb{E}[\lambda_t]$ increases if λ_c is higher.

Observe that β parameter is also affecting the structural component $\bar{\lambda}_\infty^S$ and its impact in terms of elasticity on this component is equivalent to the one produced by λ_c : indeed, Table 3 shows that the elasticities $\eta_{(\bar{\lambda}_\infty^S, \lambda_c)}$ and $\eta_{(\bar{\lambda}_\infty^S, \beta)}$ are both equal to 1. Besides this, the effect of β on the mean intensity level $\bar{\lambda}_\infty^E$ is not equivalent to the one produced by λ_c since β is affecting $\bar{\lambda}_\infty^E$ also via its

influence on the contagion component $\bar{\lambda}_\infty^C$ due to the self-excited jumps.

Reference model	Profile of trajectory	Mean intensity level $\mathbb{E}[\lambda_t]$:			Time to reach stationarity
		short-term	mid-term	long-term	
$\lambda_0 \uparrow$	modified	$\uparrow$	$\simeq$	$=$	$\simeq$
$\lambda_c \uparrow$	unchanged	$\uparrow$	$\uparrow$	$\uparrow$	$\simeq$
$\beta \uparrow$	modified	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
$\gamma \uparrow$	modified	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
$\mu \uparrow$	unchanged	$\uparrow$	$\uparrow$	$\uparrow$	$\simeq$
$\sigma \uparrow$	modified	$\uparrow$	$\simeq$	$\downarrow$	$\uparrow$
$B \uparrow$	unchanged	$\downarrow$	$\downarrow$	$\downarrow$	$\simeq$
$\delta \uparrow$	unchanged	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$

Table 4: **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 symbol “ $\uparrow$ ” corresponds to a positive impact (increase) on the quantity of interest (as compared to the reference case, see Figure 4 left side); the symbol “ $\downarrow$ ” stands for a decrease in $\mathbb{E}[\lambda_t]$; the symbol “ $\simeq$ ” means that there is roughly no change, i.e. the impact is quite negligible, while “ $=$ ” means no change at all.

5 Application to risk management

Due to new standards coming from the recent regulation, estimating key risk measures has become 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 to estimate the lapse risk in practice, i.e. the S-shaped curve, the Solvency II standard formula, and the proposed 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 statistical estimation of the parameters involved in the proposed model. How can these parameters be estimated from empirical data by an insurance company? Consider firstly the initial intensity level λ_0 and its long-run stationary mean λ_c . As already discussed in Section 2, λ_0 represents the initial force of lapse and is constant: the underlying lifetime distribution before lapsing is thus exponential. Therefore, for a given portfolio and using a moment-based estimation, a suitable value for λ_0 could be the inverse of the empirical mean of *times before lapse* in a stable regime. Of course, this estimator is not efficient and presents some limitations, contrary to the maximum likelihood estimator which has nice asymptotic properties. However it has the advantage to be simple and easy-to-implement, whereas the likelihood approach would require to derive the expression of the likelihood of the model. Approximation formulae (e.g. Taylor series) or parameter inference for the likelihood of such a model seems quite complex and is beyond the scope of this paper, but can surely be addressed in a further research project by extending existing works (e.g. [33], [23], and references therein).

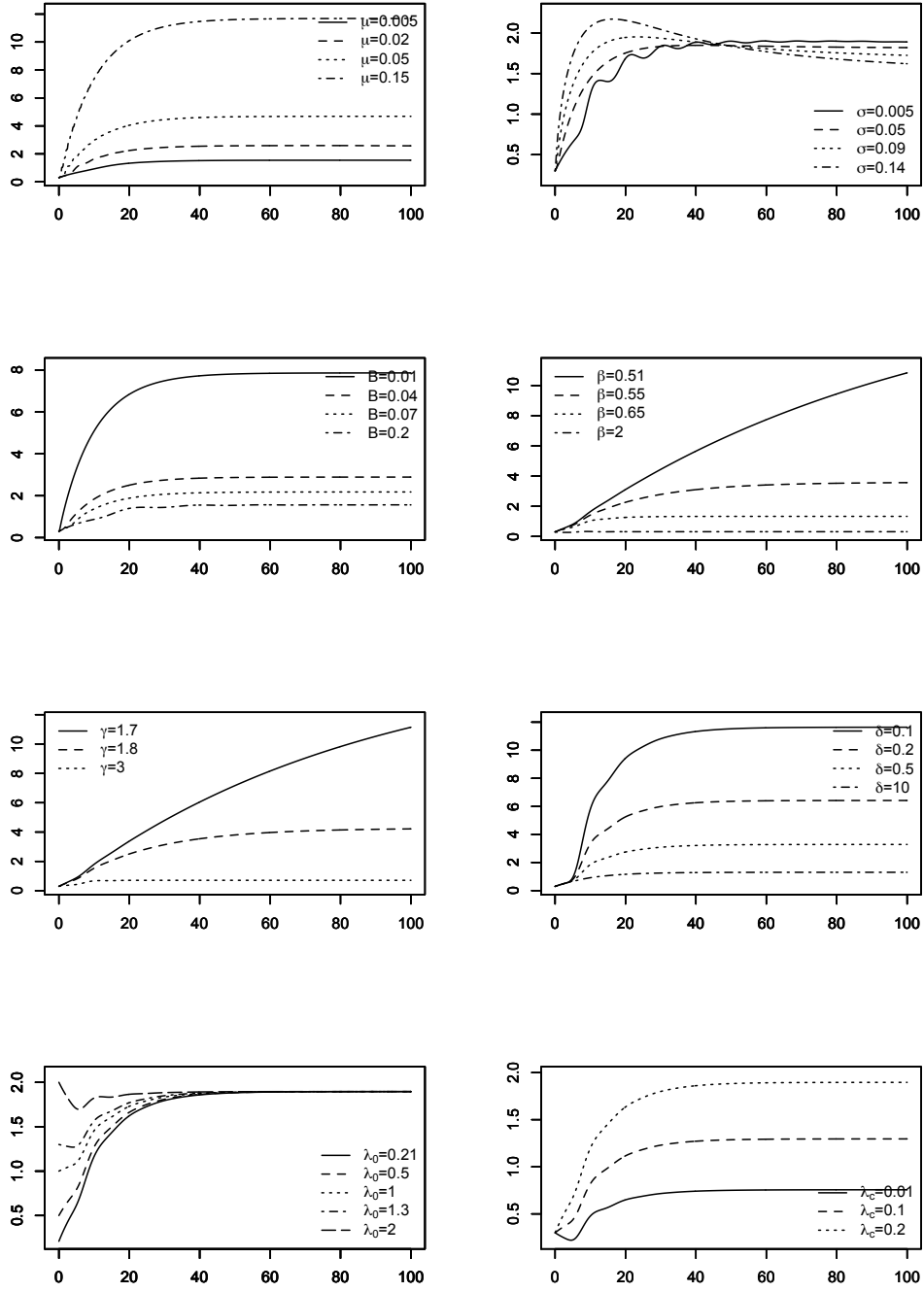


Figure 5: **Sensitivity of the expected lapse intensity $\mathbb{E}[\lambda_t]$ to the model parameters for $t \in [0, 100]$.** The comparison is made referring to the case depicted on the *left* in Figure 4, via a coeteris paribus analysis. Everything else fixed, the sensitivity is analyzed by changing only the value of the parameter under study.

Inside an insurance company, the empirical mean of *times before lapse* 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), we should experience around 75 lapses (the yearly structural lapse rate would then equal 7.5%). With a similar logic, the long-run lapse intensity level λ_c can be fixed by the risk managers as the target level embedding limits coming from the risk management guidelines. 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 $\lambda_c = 0.2$.

Other parameters are more or less connected to the risk appetite of the company. For instance, β relates to the ability of the company to reassure the policyholders, and γ, δ tie in with the mean size (in terms of sum insured) of lapsed contracts. Typically, δ should be lower than γ , 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 estimation of δ and γ is thus directly linked to the wealth of policyholders and depends on the distribution of the intensity jump sizes. The parameters of the interest rate dynamics can be estimated from historical data; while the barrier B , capturing the sensitivity of policyholders to market movements, usually comes from an expert-based judgment (depending on companies, it can vary between 10% and 50%).

We now compare some key risk indicators, estimated through the proposed dynamic contagion model and alternative current market practices. More precisely, let us consider 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 (e.g. 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 $p \in (0, 1)$ by

$$VaR_\alpha(N_t) = \inf\{k : \mathbb{P}(N_t \leq k) \geq \alpha\}, \quad TVaR_\alpha(N_t) = \frac{1}{1-\alpha} \int_\alpha^1 VaR_t(N_t) dt.$$

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 α . It gives the information about how the SCR is modified by the consideration of both correlation and contagion effects embedded in policyholders' behaviors. Numerical results are summarized in Table 5. The first conclusion is straightforward: it seems that classical stress tests procedures (e.g. Solvency II and S-shaped approaches) bring to an overall underestimation of the real lapse risk faced by life insurance companies. This is

	Solvency II Standard formula		S-shaped curve (ONC)		Hawkes counting process			Dynamic contagion process			
Parameters	Risk level	Shocks	Risk level	Shocks	$\mathbb{E}[N_t]$	VaR_α	$TVaR_\alpha$	$\mathbb{E}[N_t]$	VaR_α	$TVaR_\alpha$	
B	10%							455	1028	1142	
	30%	75	112 ⁴	75	375 ⁵	291	776	837	312	818	930
	50%							293	778	886	
δ	0.1							2461	4286	4559	
	0.5	75	112	75	375	291	776	837	702	1460	1594
	1.5							455	1028	1142	

Table 5: Impact of contagion and correlation on $VaR_\alpha(N_t)$, $TVaR_\alpha(N_t)$ at level $\alpha = 99.5\%$, in a 1-year time horizon ($t = 250$) and for various levels of the barrier B and mean external jump size $1/\delta$. Other parameters remain the same as in Figure 4 (left).

not surprising, since correlation and contagion effects among customers behaviors are not taken into account in such methodologies. This means that in practice, the resulting estimation is not conservative and a greater reserve should instead be considered to cover unexpected potential losses coming from lapses. In this view, notice that the standard Hawkes process [17] provides lower bounds for the number of lapses in the dynamic contagion case (which is consistent with their economic and financial interpretation). Besides this, recall that the Solvency II shock applied to the baseline risk was calibrated on a single product in a specific market, and is thus clearly neither consistent nor realistic (even though it serves as a benchmark today) in all contexts. Concerning the policyholders' sensitivity to interest rate movements, Table 5 and Figure 6 exhibit a non-linear behavior. Starting from an exogenous level $B = 50\%$, a 20%-decrease of the barrier causes a 7%-increase of the expected number of lapses, whereas the latter rises by around 30% when B goes from 30% to 10%. Moreover, the $TVaR$ results show that the right tail of the distribution of the number of lapses is heavier in our setting than in all other cases considered in the comparison, including the Hawkes model.

The introduction of correlation through the *external* component has a limited impact on the quantities of interest *only* in case of a small mean size of external jumps. 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). As soon as the mean size of external jumps increases, the relationship between both quantities becomes instead exponential (orange and red areas), highlighting the materiality of this component. Intuitively, this means that the composition of the portfolio is crucial, and insurers most exposed to the lapse risk are the ones having a great proportion of rich policyholders, since they are the first-to-react with possibly huge amount of notional associated to the surrenders. The proportionality coefficient between the Solvency II shock and the one given by our dynamic contagion process can go up to 40 in our example, corroborating the results discussed above. In terms of capital requirements, one should approximately apply the same coefficient to the current estimation based on Solvency II approach in order to be conservative and cover in a proper way an extreme scenario concerning the lapse risk faced by the company.

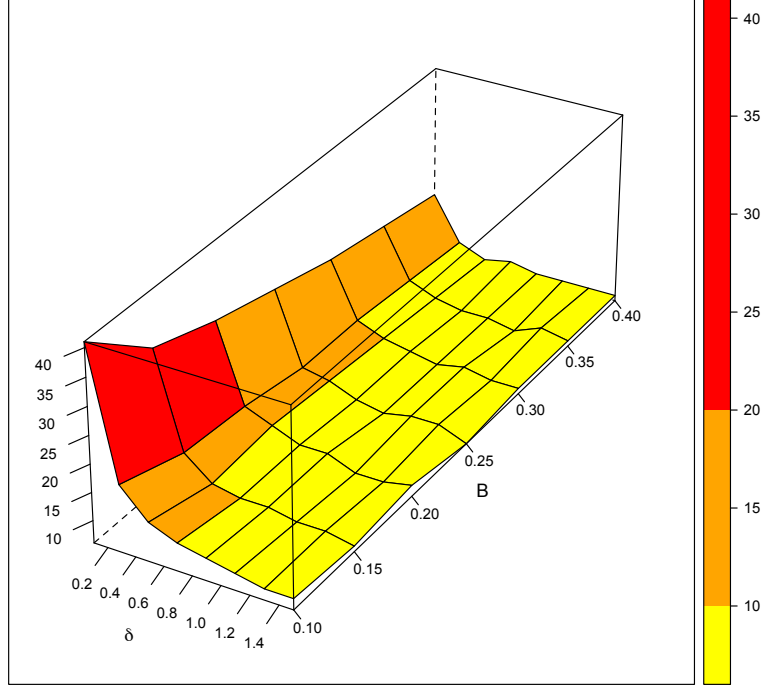


Figure 6: **Impact of B and the mean size of external jumps $1/\delta$ on the variation of the Value-at-Risk.** The z-axis is expressed as a proportionality coefficient. The comparison is made between the Solvency II shock and the one computed from the dynamic contagion process.

6 Conclusion

This paper defines a methodological proposal to dynamically integrate correlation and contagion effects among policyholders' behaviors into the modeling of lapse risk in life insurance. Indeed, correlation and contagion effects due to reactions to some '*external*' economic disturbance are crucial for companies in order to define proper stress test scenarios. This paper introduces a stochastic Hawkes-based lapse intensity: the *external* jumps affecting this intensity are assumed to be linked to a *market-driven* component, i.e. the spread between the market interest rates and the contractual crediting rate. We give in closed-form the moments of the intensity process, and provide analytic expressions of the limiting expected lapse intensity and its sensitivities to the model's parameters showing that each source of risk (*structural*, *market*, and *contagion*) has its own contribution to the long-run expected lapse level. Numerical comparisons and empirical percentiles indicate that the lapse risk margins are probably underestimated when using current practices. Possible improvements of the proposed setting could focus on separating the impacts of a global crisis scenario affecting all policyholders and a critical scenario only affecting some of them. This would allow us to set up a stress testing program able to monitor the behavioral risk more precisely. In the same spirit, risk factors affecting the structural lapse rate - such as the policyholder's age - could be integrated using a standard regression approach to model λ_0 and λ_c .

This way, the link between mortality and lapses would be more deeply taken into account ([19], [1]). Also, an alternative definition of the barrier B embedding time dependency should allow to capture the fact that policyholders' sensitivity fluctuates over time. Finally, let us mention that the characterization of the intensity process as well as the distribution of the number of lapses is of paramount importance: this could be done using the Laplace transform, or be based on a Taylor series decomposition. However, this goes beyond the scope of this paper and we leave the investigation of such interesting questions for future research.

Acknowledgments

This work is supported by both the BNP Paribas Cardif Chair “Data Analytics & Models for Insurance” and the ANR Research Project ANR-08-BLAN-0314-01. The authors would like to thank the anonymous referees for their constructive comments that contributed to significantly improve the paper.

References

- [1] Albert F.S., Bragg D.G.W. and Bragg J.M. (1999): *Mortality Rates as a function of Lapse Rates*, Actuarial research clearing house, Vol. 1.
- [2] Brace A., Gatarek D. and Musiela M. (1997): *The Market Model of Interest Rate Dynamics*, Mathematical Finance, 7(2), 127-154.
- [3] Brémaud P. and Massoulié L. (2002): *Power Spectra of General Shot Noises and Hawkes Point Processes with a Random Excitation*, Advances in Applied Probability, Vol. 34:1, p. 205–222.
- [4] Brigo D., Mercurio F. (2006): *Interest rate models: theory and practice. With smile, inflation and credit*, Springer.
- [5] Buchardt K. and Moller T. and Schmidt K. B. (2015): *Cash Flows and Policyholder Behaviour in the semi-Markov life insurance setup*, Scandinavian Actuarial Journal, Vol . 8, p.1–29.
- [6] Carver L. (2011): *Estimating and reducing lapse risk*, Insurance Risk, Feb 2011.
- [7] Dar A. and Dodds C. (1989): *Interest Rates, the Emergency Fund Hypothesis and Saving through Endowment Policies: Some Empirical Evidence for the U.K.*, Journal of Risk and Insurance.
- [8] Dassios A. and Jang J.-W. (2003): *Pricing of Catastrophe Reinsurance and Derivatives Using the Cox Process with Shot Noise Intensity*, Finance & Stochastics, Vol. 7:1, p. 73–95.
- [9] Dassios A. and Jang J.-W. (2005): *Kalman-Bucy Filtering for Linear System Driven by the Cox Process with Shot Noise Intensity and its Application to the Pricing of Reinsurance Contracts*, Journal of Applied Probability, 42:1, p. 93–107.
- [10] Dassios A. and Jang J.-W. (2013): *A bivariate shot noise self-exciting process for insurance*, Insurance: Mathematics and Economics, Vol. 53:3, p. 524–532.

- [11] Dassios A. and Zhao H. (2011): *A dynamic contagion process*, Advances in Applied Probability, Vol. 43:3, p. 814–846.
- [12] Davis M.H.A. (1984): *Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models*, Journal of the Royal Statistical Society. Series B. Methodological, Vol. 46:3, p. 353–388.
- [13] Dutang C. and Albrecher H. and Loisel S. (2013): *Competition between non-life insurers under Solvency constraints: a game-theoretic approach*, European Journal of Operational Research, Vol. 231, n. 3.
- [14] European Banking Authority (2015): *Guidelines on the management of interest rate risk arising from non-trading activities*, Final report (May 2015).
- [15] Embrechts P. and Schmidli H. and Grandell J. (1993): *Finite-time Lundberg inequalities in the Cox case*, Scandinavian Actuarial Journal, n. 1, p. 17–41.
- [16] Errais E. and Giesecke K. and Goldberg L.R. (2010): *Affine Point Processes and Portfolio Credit Risk*, Siam Journal of Financial Mathematics, Vol. 1, p. 642–665.
- [17] Hawkes A.G. and Oakes D. (1974): *A cluster process representation of a self-exciting process*, Journal of Applied Probability, p. 493–503.
- [18] Jamishidian F. (1996): *LIBOR and swap market models and measures*, Finance & Stochastics, 1, p. 293–330.
- [19] Jones B.L. (1998): *A model for analyzing the impact of selective lapsation on mortality*, North American Actuarial Journal, 1(2).
- [20] Kagraoka Y. (2005): *Modeling Insurance Surrenders by the Negative Binomial Model*, Working Paper, Musashi University (Japan)
- [21] Kim C. (2005): *Modeling Surrender and Lapse Rates with Economic Variables*, North American Actuarial Journal, p. 56–70.
- [22] Kuo W., Tsai C., and Chen W.-T. (2003): *An Empirical Study on the Lapse Rate: the Co-integration Approach*, Journal of Risk and Insurance, 70:3, p. 489–508.
- [23] Lee Y., Lim K.W., and Ong C.S. (2016): *Hawkes Processes with Stochastic Excitations*, Proceedings of the 33rd International Conference on Machine Learning, New York. JMLR: W&CP Vol. 48.
- [24] Léveillé G. and Garrido J. and Fang Wang Y. (2010): *Moment generating functions of compound renewal sums with discounted claims*, Scandinavian Actuarial Journal, n. 3, p. 165–184.
- [25] Loisel S. and Milhaud X. (2011): *From deterministic to stochastic surrender risk models: impact of correlation crises on economic capital*, European Journal of Operational Research, Vol. 214:2, p. 348–357.
- [26] Loisel S. and Arnal P. and Durand R. (2010): *Correlation crises in insurance and finance, and the need for dynamic risk maps in ORSA*, Working paper, hal-00502848.

- [27] Milhaud X. (2013): *Exogenous and endogenous risk factors management to predict surrender behaviours*, ASTIN Bulletin, Vol. 43:3, p. 373–398.
- [28] Milhaud X. and Maume-Deschamps V. and Loisel S. (2013): *Surrender triggers in Life Insurance: what main features affect the surrender behavior in a classical economic context?*, Bulletin Français d’Actuariat, 22, p. 5–48.
- [29] Miltersen K., Sandmann K. et Sondermann D. (1997): *Closed Form Solutions for Term Structure Derivates with Log-Normal Interest Rates*, Journal of Finance, 52(1), p. 409–430.
- [30] Musiela M. and Rutkowski M. (1997a): *Martingale Methods in Financial Modelling*, Springer, Berlin Heidelberg New York.
- [31] Musiela M. and Rutkowski M. (1997b): *Continuous time term structure models: Forward measure approach*, Finance & Stochastics, 1, p. 261–291.
- [32] Outreville J.-F. (1990): *Whole-life insurance lapse rates and the emergency fund hypothesis*, Insurance: Mathematics and Economics, Vol. 9, p. 249–255.
- [33] Ozaki T. (1979), *Maximum likelihood estimation of Hawkes’ self-exciting point processes*, Ann. Inst. Statist. Math., Vol. 31, no. 1, p. 145–155.
- [34] Renshaw A. E. and Haberman S. (1986): *Statistical analysis of life assurance lapses*, Journal of the Institute of Actuaries, Vol. 113, p. 459–497.
- [35] Rutkowski M. (1999): *Models of forward LIBOR and swap rates*, Applied Mathematical Finance, 6, p. 29–60.
- [36] Schwarz W. (2002): *On the convolution of inverse Gaussian and exponential random variables*, Communications in Statistics - Theory and Methods, Vol. 31:12, p. 2113–2121.
- [37] Serfozo R. (2009): *Renewal and Regenerative Processes*, In *Basics of Applied Stochastic Processes*, Springer, p. 99–167.
- [38] Seshadri V. (1999): *The inverse Gaussian distribution*, Springer.
- [39] Shreve S.E. and Karatzas I. (1991): *Brownian motion and stochastic calculus*, Springer.
- [40] Society of Actuaries (2014): *Modeling of Policyholder Behavior of Life insurance and Annuity Products. A survey and literature review*, Report.

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

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

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

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

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 [24] 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 [24] 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 (14) and the consecutive derivation. First, we have

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

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

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

Thus, by induction, some algebraic calculus and the relationship linking the $I_k(t, \theta)$ and $\widehat{I}_k(t, \theta)$ to their consecutive derivatives, we come with the desired recursion of the $m^{(n)}(t, \theta)$. ■

A.4 Proof of Lemma 2

(i) Letting $n = 1$ in Equation (19), gives

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

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

$$m^{(1)}(t, 0) = (\lambda_c + (\lambda_0 - \lambda_c)e^{-\beta t}) + e^{-\beta t} \xi^{(1)}(t, 0) + e^{-\beta t} \widehat{\xi}^{(1)}(t, 0). \quad (31)$$

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 (17)

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

which makes (31) be an integral equation. Thus, it suffices to substitute $\xi^{(1)}(t, 0)$ and $\widehat{\xi}^{(1)}(t, 0)$ (given in (18)) into (31) 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 (22) can be interpreted as the convolution of inverse Gaussian and exponential random variables, see [36]. By defining $k := (\theta_2/\theta_1^2) - 2(\beta - \frac{1}{\gamma})$, where (θ_1, θ_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 first moment of the lapse intensity strongly depends on the limit of the last term in (20). The latter is nothing but the convolution of the derivative h' of the renewal function with the exponential function, with $h(t)$ given in (12). Using the key renewal theorem, see [37, 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 (22) tends to the limit given in (25) as t goes to $+\infty$. ■