FAST CHANGE DETECTION ON PROPORTIONAL TWO-POPULATION HAZARD RATES

NICOLE EL KAROUI, STÉPHANE LOISEL, YAHIA SALHI AND CHRISTIAN MAZZA

Abstract. We consider the problem of detecting an abrupt change on the structural relationship between two mortality rates. In this paper, we are particularly interested in a form of relationship between two population mortality based on the so-called proportional hazard rate framework. In order to set up an optimal detection procedure, we broadly follow the framework introduced in Lorden (1971), adapted to the case of sequentially observed Poisson process. In this case, we show that the cumulative sums algorithm is still optimal in the case of inhomogeneous Poisson process but for criterion that replaces the expected delay by the number of events until detection. To do so, an interesting connection is made between the initial problem and the classical ruin theory. Finally, a numerical analysis is provided based on Monte Carlo simulations but also on real datasets of insured and national populations.

1. Introduction

This paper explores a monitoring scheme for Cox-like two-population model as to account for possible future breakpoints. We want to detect abrupt changes in the relationship proposed by Cox (1972) between the mortality intensities of two populations. Numerous detection approaches are available in the literature, see Page (1954), Shiryaev (1963), Roberts (1966), Siegmund (1985) in the discrete-time and Kailath and Poor (1998) for a detailed review. In the continuous-time we refer to Moustakides (2004) and the references therein. As we do not hold enough data on the occurrences of changes and no priors are available, we follow the formulation introduced by Lorden (1971). We consider that the change occurs at some unknown but deterministic time. Moreover, we assume that the pre-change parameter is an input of the procedure, which may stem from experts opinions.

For the Lorden's framework, the CUSUM test (procedure), initially proposed by Page (1954), is generally used as a means to detect sequentially changes in the distribution of random processes. The optimality of the CUSUM test has been established for different cases, both on discrete and continuous time. For example, for independent and identically distributed variables with known distribution

Received by the editors April 16, 2013.

1991 *Mathematics Subject Classification.* Primary 62L15; Secondary 97M30; 62P05.

Key words and phrases. Proportional Hazard, Change-Point, Sequential Detection, Poisson Process, Longevity Risk.

before and after the change, Lorden (1971) shows that the CUSUM statistics corresponds to the optimal detection strategy. Similarly, Ritov (1990) use a Bayesian property of the CUSUM statistics and provide an alternative proof of optimality in Lorden's sense. Later on, Moustakides (1998) have extended the optimality results to a special class of dependent processes. Optimal CUSUM procedures were proposed by Poor (1998) for exponentially penalized detection delays. In continuous time, the optimality of CUSUM has been established for Brownian motion with constant drift by Beibel (1996), using the Bayesian setting of Ritov (1990), which also yielded optimality in Lorden's sense, and by Shiryaev (1996). Recently, an extension of the optimality for the CUSUM test is established by Moustakides (2004) to detect changes in Itô processes when the expected delay is replaced with the corresponding relative entropy in the criterion. Finally, Fellouris and Chronopoulou (2013) establish the optimality of the CUSUM test with respect to a modified version of Lorden's criterion for arbitrary processes with continuous paths and apply this general result to the special case of fractional diffusiontype processes.

In this paper, we follow broadly the framework of Lorden (1971) adapted to the case of point processes, with a given time-dependent intensity. In the same vein as Shiryaev (1996), Beibel (1996), Moustakides (2004) and Fellouris and Chronopoulou (2013), we show that the so-called cumulative sums (CUSUM) stopping time is still optimal when one wants to minimize the average number of events until detection (instead of the average delay before detection). Furthermore, using an interesting connection of the CUSUM process with ruin theory, we propose explicit formulas for the performance measures, respectively the worst detection delay and the average run length. Aside from the theoretical result on the optimality, this framework is also interesting from the practical point of view due to the large use of the Cox's regression model in the life actuarial fields especially for experienced mortality forecasting.

The remainder is organized as follows. First, in Section 2, we set up the notation and terminology that shall be useful throughout the paper. Additionally, we introduce the uncertainty surrounding our modelling approach being related to proportional hazard rate framework. In Section 3, we introduce the quickest detection problem using a time to event performance measure. The optimality of the proposed procedure is also demonstrated as well as the computation of performance measures. Finally, an illustration is given in Section 4. First, we draw a numerical analysis based on Monte Carlo simulation to investigate the efficiency of the algorithm. Next, we apply the methodology to two real-world England and Wales mortality datasets.

2. Notation and Formulation of the Problem

2.1. **Proportional Hazard Rate Model Uncertainty.** On a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, satisfying the usual conditions of right continuity and completeness, we consider an inhomogeneous Poisson process $N = (N_t)_{t\geq 0}$ counting the deaths within a population, which may either correspond to some age group, cohort or merely to the whole portfolio.

Denote by $\Lambda = (\Lambda_t)_{t\geq 0}$ the compensator of N, defined as the unique $\mathbb{F}$-predictable increasing process such that the compensated process $M = N - \Lambda$ is a martingale. In the sequel, we assume that Λ is absolutely continuous with respect to the Lebesgue measure in such a way that there exists a non-negative $\mathbb{F}$-predictable process $\lambda = (\lambda_t)_{t\geq 0}$ satisfying $\Lambda_t = \int_0^t \lambda_s ds$ for $t \geq 0$, see Brémaud (1981, pp. 27-29) for more details. In accordance to the survival analysis literature, we call Λ the martingale hazard process and λ the intensity. This approach is simple but consistent with the well-known Poisson log-bilinear model of Brouhns et al. (2002), where intensity λ is given by the instantaneous exposure to death multiplied by the force of mortality, see .

Unlike the common use of such a framework, we only focus on the case of deterministic evolution of mortality. That is, the process λ is deterministic, which may not be restrictive from a practical point of view. Indeed, mortality models have traditionally been based on deterministic life tables that allow for mortality improvements[1]. However, one may rely on stochastic models (see for example Cairns et al. (2009) and reference therein), so the intensity in our case may be seen as an average scenario. Most of these methodologies are well suited for large populations, e.g. national populations, whose mortality can be decomposed into different factors with less bias in the parameter estimation. Thus, these models are not directly applicable to insurance portfolios and one may only rely on adjustment to a reference mortality. Therefore, practitioners first determine a forecast for a reference mortality and hence adjust the insurance portfolio to the latter. This is generally done using the so-called relational models, which allow to link demographic indicators in two groups of individuals, e.g. mortality intensities. Among those frameworks we can refer to Brass (1971) and Cox (1972). The latter proposed a proportional hazards framework, which assumes that the mortality intensity of a group of interest is proportional to a reference population. The proportionality factor does not depend on age and is constant all over the projection period. More specifically, the mortality intensity λ_t is linked to a reference rate λ_t^0 as follows

$$\lambda_t = \alpha \lambda_t^0, \quad \text{for} \quad t \geq 0, \tag{2.1}$$

where α is a deterministic positive parameter. Yet the portfolio of policyholders may change and evolve over time due to product design or to underwriting policy. By change we mean the change of mortality profile of the insurance portfolio which would perturb the stability over time of the relationship in the above model. In addition, mis-estimation of the factor α turns out to be of substantial concern especially for volatile or limited dataset.

[1]This may be seen as a best estimate intensity, which represents the average evolution of mortality over time based either on a stochastic model, experts opinion or on some internal methodologies.

Therefore, it is desirable to account for model uncertainty in (2.1), as it may give rise to an uncertainty surrounding the relevance of the parameter α during the projection period. More formally, we assume that the initial level of the intensity λ with regard to the reference intensity λ^0 is given by $\underline{\alpha}$, and, at some unobservable time θ, the level may change to some value $\overline{\alpha}$.
We can now formulate the underlying model on the following form

$$\lambda_t = \alpha_t \lambda^0 \quad \text{for} \quad t \geq 0,$$

with $\alpha_t = \underline{\alpha}\mathbf{1}_{\{t<\theta\}} + \overline{\alpha}\mathbf{1}_{\{t\geq\theta\}}$, or equivalently:

$$\lambda_t = \begin{cases} \underline{\alpha}\lambda_t^0, & t < \theta, \\ \overline{\alpha}\lambda_t^0, & t \geq \theta. \end{cases} \tag{2.2}$$

Throughout the sequel, we assume that the proportional factors $\overline{\alpha}$ and $\underline{\alpha}$ are predefined and deterministic. For simplicity of notation, we let $\rho = \overline{\alpha}/\underline{\alpha}$ and without loss of generality fix $\underline{\alpha} = 1$. Furthermore, motivated by empirical findings which show that the proportional rate increases over time, we consider the case when $\rho > 1$. Of course, for a pension fund or an insurer, basis risk mainly corresponds to the opposite case, where the spread between mortality rates of the two populations (policyholders population and national population on which risk management strategies and risk quantification are often based) increases instead of decreasing. This case may be addressed with our method after either permuting the two populations or considering a detection barrier in the opposite direction. In this paper, we illustrate our method on the case $\rho > 1$ because it corresponds to what occurred previously for the two England and Wales populations for which data is available. Note that the considered case ($\rho > 1$) and the other one ($\rho < 1$) are both interesting for investors and cedants in the recent life insurance securitization set up by Swiss Re, because the payoff of this ILS product depends on the evolution of the difference between U.S. mortality and E.& W. longevity indices. More generally, large insurance and reinsurance groups with mortality exposure in one country or one region and longevity exposure in another zone must carefully check that no significant break occurs in the quotient of mortality intensities of the two populations.

Remark 1. *Another way to look at the model in (2.2) is to consider a single underlying mortality such that the break on the model may not only occur due to a change of the mortality profile but also due to shifts of the trend. Even if this particular case is not fully specified in our analysis, the results below may be directly applied and the following reasoning holds as well. It is worth pointing out that in such a case the parameter ρ stands for the slope of mortality intensity trend.*

2.2. **Probabilistic Formulation.** In what follows, we assume that the change-point θ is unknown but deterministic such that $\theta \in [0, \infty]$. We shall formulate the probabilistic model underlying the change-point model in (2.2) as follows. First, let $\mathbb{P}_\theta$ be the probability measure of the counting process N under the assumption that the disorder happened at time θ. In particular, $\mathbb{P}_\infty$ is the distribution of N under the assumption that the disorder never happened, and $\mathbb{P}_0$ is the distribution of the process N with intensity $\rho\lambda^0$. Similarly, let $\mathbb{E}_\theta(\cdot)$ denotes the expectation with respect to the probability $\mathbb{P}_\theta$. In doing so, we stated that the distribution of N, i.e. $\mathbb{P}_\theta$, changes at the unknown time θ from $\mathbb{P}_\infty$ to $\mathbb{P}_0$. Thus, $\mathbb{P}_\theta$ coincides with $\mathbb{P}_\infty$ on $\mathcal{F}_t$ for $t \in [0,\theta]$ and with $\mathbb{P}_0$ for any $t \in (\theta,\infty)$. Accordingly, the following process

$$M_t^0 = N_t - \Lambda_t \quad \text{and} \quad M_t^\infty = N_t - \rho\Lambda_t, \qquad \text{for } t \geq 0,$$

are respectively $\mathbb{P}_0$ and $\mathbb{P}_\infty$ martingales.

In order to ensure that the intensity process is well-defined in view of the problem in (2.2) and the notation above, we shall make, for every $0 \leq t < \infty$ the following assumption

$$\int_0^t \lambda_s ds < \infty, \quad \mathbb{P}_0 \text{ and } \mathbb{P}_\infty \text{ a.s.} \tag{H1}$$

Furthermore, suppose that $\mathbb{P}_0$ and $\mathbb{P}_\infty$ are equivalent on $(\Omega,\mathcal{F})$, and let $U = (U_t)_{t\geq 0}$ be the corresponding log-likelihood ratio, defined as

$$U_t = \log \left.\frac{d\mathbb{P}_0}{d\mathbb{P}_\infty}\right|_{\mathcal{F}_t}, \quad t > 0,$$

with $U_0 = 0$, or more generally for every $\theta \geq 0$

$$U_t = \log \left.\frac{d\mathbb{P}_\theta}{d\mathbb{P}_\infty}\right|_{\mathcal{F}_t}, \quad t > \theta,$$

as $\mathbb{P}_\theta$ is continuous with respect to $\mathbb{P}_\infty$. In addition, it is easy to check that U has the following form

$$U_t = \log(\rho)N_t + (1-\rho)\Lambda_t, \quad t \geq 0. \tag{2.3}$$

From (2.3) we should highlight that the process U is right-continuous with left limit. Moreover, the assumption stated in (H1) makes obvious that $\exp(U)$ is a local martingale. We denote $\Delta U_t = U_t - U_{t-}$ the jump of U at $t \geq 0$, which is nothing but $\log\rho$. For any $x \geq 0$, we introduce the following processes needed throughout the sequel:

$$L_t = -\underline{U}_t \quad L_t(x) = (L_t - x)^+. \tag{2.4}$$

The process L is the local time, which is clearly non-negative and only increases on the subsets of jumps of the underlying process U, and $L(x)$ is the reflected local time at x.

Definition 1 (Reflected Log-Likelihood Process)**.** *For $x \geq 0$, we introduce the reflected version of the process U, denoted $V(x)$, starting from x as follows*

$$V_t(x) = x + U_t + L_t(x). \tag{2.5}$$

The above process will play a crucial role in the sequel. This is nothing but the so-called cumulative sums (cusum) process introduced by Page (1954) with a head-start x as in Lucas and Crosier (1982). To see this, it suffices to write the process in the following form

$$V_t(x) = x + U_t + L_t(x) = U_t + x \vee L_t = U_t - (-x) \wedge \underline{U}_t. \tag{2.6}$$

For $x = 0$, remarking that $(-x) \wedge \underline{U}_t$ starts from 0 and becomes the running minimum of U, we recover the regular version initially introduced by Page (1954). The above definition makes the reflected process right-continuous as the paths of V are slightly different from the ones of U. More precisely, it is easily seen that $V(x)$ only jumps during the arrival times of U, i.e. $T_1, T_2, \cdots$. Indeed, in view of (2.6) we might write (2.5) as follows:

$$V_t(x) = \begin{cases} U_t + x, & \underline{U}_t \geq -x \\ U_t - \underline{U}_t, & \underline{U}_t < -x. \end{cases} \tag{2.7}$$

Thus, the process has the same behavior as U above the barrier $-x$, jumping at the same amount at times $T_1, T_2, \cdots$, with size $\log \rho$, i.e. see (2.7). In between jumps, $V(x)$, decreases at rate $(1-\rho)\lambda_t$ which is cut off when $V_t(x) = 0$, i.e. $\underline{U}_t \leq -x$. In other words, the reflection only intervenes when U hits the barrier $-x$, at which $V(x)$ is equal to 0 and remains so until the next jump of U. The latter can be seen from (2.7). These properties can be summarized in the following equation

$$V_t(x) = x + (\log \rho) N_t + (1-\rho) \int_0^t \lambda_s \mathbf{1}_{\{V_s(x)>0\}} ds. \tag{2.8}$$

In the basic case of quadratic intensity, the sample path relation, for $0 \leq x < m$, between these processes is illustrated in Figure 1.

From the construction above it is intuitively clear that $V(x)$ is Markov. This is checked as well as the strong Markov property in the following lemma, see also Asmussen (2003, Proposition 2.1, p. 251). First, let $U_t^{(T)} = U_{t+T} - U_T$, $L_t^{(T)} = -\underline{U}_t^{(T)}$ and $L_t^{(T)} = (L_t^{(T)} - x)^+$ then the following ensures that the process V is Markov.

Lemma 1. *With above notations we have* $V_{t+T}(x) = V_T(x) + U_t^{(T)} + L_t^{(T)}(V_T(x))$.

Proof. By abuse of notation, we write V_{t+T} instead of $V_{t+T}(x)$. Then, the stated expression for V_{t+T} is the same as:

$$\begin{aligned} U_t^{(T)} + V_T \vee L_t^{(T)} &= U_{t+T} - U_T + (U_T + x \vee L_T) \vee \Big(U_T - \inf_{0 \leq \nu \leq t} U_{\nu+T} \Big) \\ &= U_{t+T} + x \vee \Big(\sup_{0 \leq \nu \leq T} -U_\nu \Big) \vee \Big(\sup_{0 \leq \nu \leq t} -U_{\nu+T} \Big) \\ &= U_{t+T} + x \vee \Big(\sup_{0 \leq \nu \leq t+T} -U_\nu \Big) = V_{t+T}, \end{aligned}$$

which is the desired formula. □

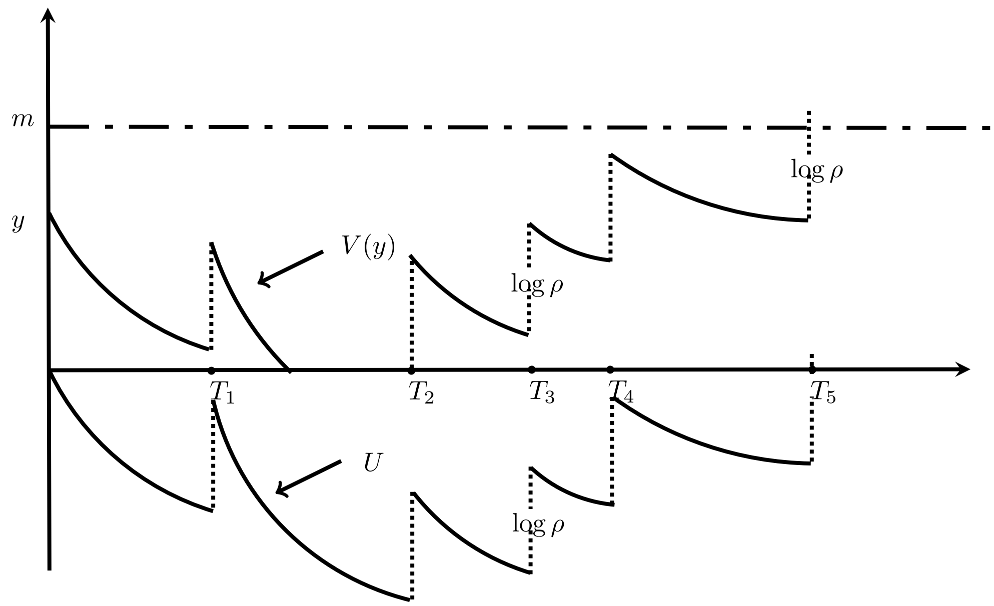

Figure 1: *Example of the path of the processes U and V. The process V coincides with U before hitting the level* 0.

The above proposition shows that V_{t+T} is constructed from $U_t^{(T)}$ and V_T in the same way as $V_T(x)$ is constructed from U_t and x. This allows to start the process V_{t+T} from the intermediate state V_t. Moreover, it is clear that V_{t+T} does not depend on the information before time t. The content of Lemma 1 is that $V(x)$ evolves as U until the first hitting time of 0, so that the crux is to merely describe the behavior starting from $V_0(x) = 0$ or, equivalently, $x = 0$. The strong Markov property immediately follows from an inspection of the paths, we refer to Asmussen (2003, Proposition 7.2, p. 109) for a detailed proof.

Given the initial value of the reflected process $V(x)$ we adopt the notation $\mathbb{P}_\theta^{(x)}$ ($\mathbb{E}_\theta^{(x)}[\cdot]$) which stands for the probability measure (respectively the expected value) conditionally on the initial value $V_0 = x$. Next, let:

$$\tau_m(x) = \inf\{t \geq 0 : V_t(x) \geq m\} \tag{2.9}$$

be the $\mathbb{F}$-stopping time defined as the first passage to level $m > 0$.

When no confusion can arise, we will simply write respectively V_t and τ_m instead of $V(0)_t$ and $\tau_m(0)$. Moreover, we shall omit the subscript in $\mathbb{P}_\theta^{(0)}$ and $\mathbb{E}_\theta^{(0)}[\cdot]$.

Having disposed of this notation, we can now return to the problem in (2.2). In the same vein as Moustakides (2004), we will need throughout the sequel assume that the process N is $\mathbb{P}_0$-a.s. non-terminating in the sense that

$$\mathbb{P}_0\left(\lim_{t\to\infty} N_t = \infty\right) = 1. \tag{H2}$$

This assumption implicitly imposes that the population under study is non-decreasing over time. From a practical point of view, this obviously excludes run-off portfolios and closed books where the scheme is closed to new policyholders entries. This may be in adequation for insurance portfolios in the build-up phase with a continuing arrival of new policyholders. In a general setting of a given population, such a hypothesis imposes the presence of births or immigration which may ensure the non-termination of the underlying population. Assumption (H2) is more appealing and seems particularly important as far as we are looking for changes on the mortality profile and mortality trend and not changes which may be induced by reduction of population under study. This would typically lead to some bias on the parameter estimate due to the limited number of records. Under (H2), we should enlarge the filtration $\mathbb{F}$ to incorporate the information coming from the births, immigration or individuals' arrival in the portfolio. Without loss of generality, we stick on the initial specification as the information is uniquely specified by the counting process. Roughly speaking, we implicitly assume that this additional information is deterministic. Another implication of (H2) is that the time horizon of the problem is by construction infinite.

3. Quickest Detection Problem

3.1. **Quickest Detection Problem Formulation.** The goal in the sequel is to detect the occurrence of the change at time θ. More precisely, we are looking for an $\mathbb{F}$-adapted stopping time τ which declares the occurrence of the change while minimizing a suitable delay criterion. We will look for stopping times belonging to the set of events $\{t \geq \theta\}$ while avoiding false detection, i.e. $\tau \in \{t < \theta\}$. To design such a procedure we will need two types of performance measures: One being a measure of the delay between the time a change occurs and the time it is detected and the other being a measure of the frequency of false alarms. Given that no priors are available on the change-point θ, we will follow the procedure proposed by Lorden (1971), which advocates penalizing the detection delay via its worst-case value. The worst case detection delay is adopted as a measure of the detection lag, conditioned on the observations before the change time θ, which is given by:

$$\sup_{\theta\in[0,\infty)} \left(\operatorname{ess\,sup} \mathbb{E}_\theta\left[(\tau-\theta)^+ | \mathcal{F}_\theta\right]\right). \tag{3.1}$$

In addition, the desire to make the above quantity small, i.e. quick detection, is balanced with a constraint on the rate of false alarms measured by the average run length (ARL) given by $\mathbb{E}_\infty[\tau]$. In our case, it is more practical to adopt the same philosophy. However, as the insurers are more used to handle actuarial quantities that have more monetary signification, especially in terms of the incurred losses, we propose a different alternative performance measure. Proceeding along the same lines in Gandy et al. (2010) we consider a *time to event* framework. More formally, we introduce the following worst-case conditional time to event performance measure:

$$R(\tau) = \sup_{\theta \in [0,\infty)} \left(\operatorname{ess\,sup} \mathbb{E}_\theta \left[(N_\tau - N_\theta)^+ | \mathcal{F}_\theta \right] \right). \tag{3.2}$$

The delay functional described above measures mortality displacement stemming from the model in (2.2) in terms of *excess of mortality*, which is well celebrated by life insurance practitioners. Consequently, the desire to monitor the detection and thus sounding an alarm *as quick as possible* is meant to find a stopping time τ which minimizes the excess of mortality between the change occurrence of change and the detection time. This gives arise to the following min-max problem:

$$\inf_\tau R(\tau) = \inf_\tau \sup_{\theta \in [0,\infty)} \left(\operatorname{ess\,sup} \mathbb{E}_\theta \left[(N_\tau - N_\theta)^+ | \mathcal{F}_\theta \right] \right). \tag{3.3}$$

In order to control the false alarm we advocate the use of the expected number of deaths until stopping, $\mathbb{E}_\infty [N_\tau]$, see also Gandy et al. (2010). We hence consider a class of $\mathbb{F}$-stopping times τ that satisfies the false alarm constraint:

$$\mathbb{E}_\infty [N_\tau] \geq \omega. \tag{3.4}$$

For continuous filtrations generated by a Browian motion, an other modification has been proposed by Moustakides (2004) and it is in the same vein as the original formulation introduced by Lorden (1971). Moustakides (2004) used the Kullback-Leibler divergence between the pre-change and post-change probability measures instead of the actual time of delay to detection a change on the drift of a Brownian motion. A more general formulation was proposed recently in Fellouris and Moustakides (2011) where detection delay and false alarm rate are measured in terms of the expected accumulated quadratic variation until the alarm. In both cases, it is shown that a test based on the reflected process in (2.5), for $x = 0$, solves the Lorden (1971)'s optimization problem, as long as the threshold level m is chosen in such a way that the false alarm constraint be satisfied with equality.

The aim of the following is to prove the optimality of the stopping time τ_m such that $\mathbb{E}_\infty[N_{\tau_m}] = \omega$ for the problem in (3.3). At this point, it is appropriate to introduce the two functions h_0 and h_∞ of x defined as:

$$h_0(x) = \mathbb{E}_0^{(x)} \left[N_{\tau_m(x)} \right] \quad \text{and} \quad h_\infty(x) = \mathbb{E}_\infty^{(x)} \left[N_{\tau_m(x)} \right], \tag{3.5}$$

which are the performance of the detection procedure introduced so far. It is then clear that $h_(0)$ and h_∞ express respectively the average number of events until detection and the average time to event in case of false detection both for the regular procedure, i.e. $x = 0$.

3.2. **Performance Evaluation.** In this subsection, we are going to present some key properties for h_0 and h_∞. We first present a lemma that states an important property regarding the monotonicity these functions.

Lemma 2. *The functions $h_0(x)$ and $h_\infty(x)$ are decreasing and uniformly bounded for $0 \leq x < m$. Moreover, $h_0(x) = h_\infty(x) = 0$ for $x \geq m$.*

Proof. First, notice that for $x \geq m$ the process V already hit the barrier at $t = 0$ and $\tau_m(x) = \inf\{t \geq 0,\ V_t(x) \geq 0\} = 0$. Thus $h_0(x) = h_\infty(x) = 0$.
For $0 \leq x < m$, by Definition 1 it is clear that $V(x)$ is increasing in x, it immediately follows that $\tau_m(x)$ is decreasing in x and so is $\mathbb{E}_0^{(x)}[N_{\tau_m(x)}]$ and $\mathbb{E}_\infty^{(x)}[N_{\tau_m(x)}]$. Consequently, we have for every $x \geq 0$

$$h_0(x) \leq h_0(0) = \mathbb{E}_0[N_{\tau_m}] \quad \text{and} \quad h_\infty(x) \leq h_\infty(0) = \mathbb{E}_\infty[N_{\tau_m}]. \tag{3.6}$$

The uniform boundedness follows immediately, which completes the proof. □

The regularity of the functions h_0 and h_∞ is established by the next theorem. We also derive partial differential equations satisfied by these functions.

Theorem 1. *The functions h_0 and h_∞ are continuously differentiable on $[0, m]$ with derivatives satisfying the following partial differential equations*

$$h_0'(x) = \frac{\rho}{1-\rho}\Big(h_0(x) - h_0(x + \log\rho) - 1\Big), \tag{3.7}$$

$$h_\infty'(x) = \frac{1}{1-\rho}\Big(h_\infty(x) - h_\infty(x + \log\rho) - 1\Big). \tag{3.8}$$

Proof. For any $\mathbb{P}_0$-a.s. finite stopping time η, by conditioning on the set $\mathcal{F}_\eta$ first and then using the strong Markov property of the process V, we have

$$h_0(x) = \mathbb{E}_0^{(x)}\left[\mathbb{E}_0^{(V_\eta)}\left[N_{\tau_m}\right]\right] = \mathbb{E}_0^{(x)}\left[h_0(V_\eta)\right]. \tag{3.9}$$

For T_1, the first jump of the process N, letting $\eta = T_1 \wedge \epsilon$ for $\epsilon > 0$ we can use classical ruin theory tools to derive the conditional expectation in (3.9). To do so, we suppose that ϵ is infinitesimal so that it may occur only one jump on the interval $[0, \epsilon]$. Indeed, the average jumps, i.e. death records, on $[0, \epsilon]$ can be approximated under $\mathbb{P}_0$ by $\rho\lambda_0\epsilon + o(\epsilon)$, and thus the probability that more than one jump appear in this interval is small enough to be ignored.
Distinguishing two cases, η being equal to T_1 and ϵ, respectively $T_1 < \epsilon$ and $T_1 > \epsilon$ we can develop the expectation in (3.9). Notice that when $T_1 < \epsilon$, h_0 will starts from the current value of $V(x)_t$ given that a jumps has occurred, i.e. $1 + h_(x + \log\rho + (1-\rho)\Lambda_t)$. Note also that conditioning on $V_0 = x$, the conditional probability of the event $T_1 > \epsilon$ is given by $e^{-\Lambda_\epsilon}$ and the conditional density of T_1 is $\lambda_t e^{-\Lambda_t}$ for $t \leq \epsilon$. Based on this heuristic argument, conditioning on the above cases, leads to

$$\begin{aligned} h_0(x) &= \mathbb{P}_0^{(x)}(T_1 > \epsilon)\, h_0(x + (1-\rho)\Lambda_\epsilon) \\ &\quad + \int_0^\epsilon \rho\lambda_t e^{-\rho\Lambda_t} h_0(x + \log\rho + (1-\rho)\Lambda_t)dt + \int_0^\epsilon \rho\lambda_t e^{-\rho\Lambda_t} dt \\ &= e^{-\Lambda_\epsilon} h_0(x + (1-\rho)\Lambda_\epsilon) \\ &\quad + \int_0^\epsilon \rho\lambda_t e^{-\rho\Lambda_t} h_0(x + \log\rho + (1-\rho)\Lambda_t)dt + \int_0^\epsilon \rho\lambda_t e^{-\rho\Lambda_t} dt, \end{aligned}$$

which is owed to the fact that jumps of the process V coincide with those of N.

Additionally, the function Λ is a bijection (due to assumption (H2)) and both Λ and its inverse, denoted Λ^{-1} are continuous. Then making the change of the time scale $\varepsilon \to \Lambda_\epsilon$ in the last equation yields to

$$h_0(x) = e^{-\rho\varepsilon} h_0(x+(1-\rho)\varepsilon) + \rho \int_0^{\Lambda_\varepsilon^{-1}} \lambda_t e^{-\rho\Lambda_t} h_0(x+(1-\rho)\Lambda_t + \log\rho)dt + \int_0^{\Lambda_\varepsilon^{-1}} \rho\lambda_t e^{-\rho\Lambda_t} dt. \tag{3.10}$$

Rearranging the terms in the above equation gives

$$h_0(x+(1-\rho)\varepsilon) = e^{\rho\varepsilon} h_0(x) - \rho e^{\rho\varepsilon} \int_0^{\Lambda_\varepsilon^{-1}} \lambda_t e^{-\rho\Lambda_t} h_0(x+(1-\rho)\Lambda_t + \log\rho)dt - e^{\rho\varepsilon} \int_0^{\Lambda_\varepsilon^{-1}} \rho\lambda_t e^{-\rho\Lambda_t} dt.$$

Letting $\varepsilon \to 0$, from the fact that $\lim_{\varepsilon\to 0} \Lambda_\varepsilon^{-1} = 0$, we see that $\lim_{\varepsilon\to 0} h_0(x+(1-\rho)\varepsilon) = h_0(x)$. Therefore, as $\rho < 1$, it follows that h_0 is right-continuous. Similarly, we proceed to show the left-continuity of h_0. To this aim, It suffices to consider (3.10) at the point $x-(1-\rho)\varepsilon$ together with a change of variable $s = (1-\rho)(\Lambda_t - \varepsilon)$ in the first integral, which gives rise to

$$h_0(x-(1-\rho)\varepsilon) = e^{-\rho\varepsilon} h_0(x) - \rho(1-\rho) \int_0^{-(1-\rho)\varepsilon} \frac{\lambda_{\Lambda^{-1}\left(\frac{s}{(1-\rho)}+\varepsilon\right)}}{\lambda_{\frac{s}{(1-\rho)}+\varepsilon}} e^{-\rho\left(\frac{s}{(1-\rho)}+\varepsilon\right)} h_0(x+\log\rho+s)ds + \int_0^{\Lambda_\varepsilon^{-1}} \rho\lambda_t e^{-\rho\Lambda_t} dt.$$

It is easy to check that the first and the last terms converge respectively to $h_0(x)$ and 0 as $\varepsilon \to 0$. The second term has a continuous and positive integrand and thus bounded locally around 0. It follows easily that this the integral also converges to 0, which shows that h_0 is right-continuous. Accordingly h_0 is continuous for each $x < m$.

We now turn to the differentiability of h_0. By subtracting $h_0(x-(1-\rho)\varepsilon)$ into (3.10), dividing $(1-\rho)\varepsilon$ on both sides and rearranging the terms, we obtain

$$\frac{h_0(x-(1-\rho)\varepsilon) - h_0(x)}{(1-\rho)\varepsilon} = \frac{\rho}{1-\rho} \left\{ \left(\frac{1-e^{\rho\varepsilon}}{\rho\varepsilon}\right) h_0(x+(1-\rho)\varepsilon) - \frac{1}{\varepsilon} \int_0^{\Lambda_\varepsilon^{-1}} \lambda_t e^{-\rho\Lambda_t} h_0(x+(1-\rho)\Lambda_t+\log\rho)dt - \int_0^{\Lambda_\varepsilon^{-1}} \rho\lambda_t e^{-\rho\Lambda_t} dt. \right\}$$

A passage to the limit when $\varepsilon \to 0$, using similar arguments as above and the (left)-continuity of h_0, implies that

$$h_0'(x) = \lim_{\varepsilon\to 0} \frac{h_0(x-(1-\rho)\varepsilon) - h_0(x)}{(1-\rho)\varepsilon} = \frac{\rho}{1-\rho}\Big(h_0(x) - h_0(x+\log\rho) - 1 \Big)$$

Here, by abuse of notation h' denotes the left derivative. Similar arguments as in the case of h_0 applies to show that a right derivative also exists and coincides with the left derivative. Finally, the continuity of h_0' immediately follows from the continuity of h_0.
The desired results concerning h_∞ can be proved using a similar analysis. □

Remark 2. *In the proof of Theorem 1, we deliberately omit the case when U hits the reflecting barrier $-x$ and thus V vanishes. We implicitly assumed that one can choose an infinitesimal interval $[0, \epsilon$ in such a way that the process remains above this barrier. By doing so, we only need to assume that $\epsilon < \Lambda^{-1}_{\frac{x}{\rho-1}}$.*

Remark 3. *In order to solve the partial differential equations (3.7) and (3.8), we will need the boundary condition $h_0(m) = h_\infty(m) = 0$, together with the boundedness property imposed by Lemma 2.*

To derive explicite formula for h_0 and h_∞ we can, at first, attempt to solve the p.d.e. in Theorem 1. However, It is worth mentioning one can use direct method using some known results in risk theory. First, consider a random change of time in (3.5) of the form $\tilde{\tau}_m(x) = \Lambda_{\tau_m(x)}$, made possible due to assumption (H2). Indeed, as a consequence of (H2) we have $\lim_{t\to\infty} \Lambda_t = \infty$, $\mathbb{P}_0^{(x)}$-a.s. Thus the time change defines $\tau_m(x)$ for all $\tilde{\tau}_m(x)$. Hence, denote $\tilde{N}$ the counting process given by $\tilde{N}_{\tilde{\tau}_m(x)} = N_{\tau_m(x)}$ being a standard Poisson process with intensity 1 (see Brémaud (1981, Theorem 16, p. 41)). For any integer $n \in \mathbb{N}$, note that the stopping time $\tilde{\tau}_m^n(x) = \tilde{\tau}_m(x) \wedge n$ is $\mathbb{P}_0^{(x)}$-a.s. bounded, and $\tilde{N}_{\tilde{\tau}_m^n(x)} - \tilde{\tau}_m^n(x)$ is a $\mathbb{P}_0^{(x)}$-martingale. By optional stopping, it follows that

$$\mathbb{E}_0^{(x)}[\tilde{N}_{\tilde{\tau}_m^n(x)}] = \mathbb{E}_0^{(x)}[\tilde{\tau}_m^n(x)].$$

Letting $n \to \infty$, we have $\tilde{\tau}_m^n(x) \to \tilde{\tau}_m(x)$ monotonically. In the previous equality, using the monotone convergence we obtain that $\mathbb{E}_0^{(x)}[\tilde{N}_{\tilde{\tau}_m(x)}] = \mathbb{E}_0^{(x)}[\tilde{\tau}_m(x)]$, being equal to $\mathbb{E}_0^{(x)}[N_{\tau_m(x)}]$.

Corollary 1. *h_0' and h_∞' are uniformly bounded in $[0, m)$*

Proof. On account of Theorem 1, we have for each $x < m$

$$|h_0'(x)| = \frac{\rho}{\rho-1}\big|h_0(x) - h_0(x + \log\rho) - 1\big| \leq \frac{\rho}{\rho-1}\big(2h_0(0) + 1\big).$$

Thus h_0' is uniformly bounded on $[0, m)$. Likewise, one can show the uniform boundedness of h_∞'. □

3.3. **Maximal Inequalities and Optimal Stopping.** The problem of the form (3.3) is widely discussed in the literature in many applied contexts including in sequential detection problems. This is generally solved through maximal inequalities as shown in Graversen and Peškir (1998). As we may show later in the case of non-homogeneous observed Poisson process we derive similar results allowing to state the solution of optimality. More precisely, we show that τ_m is optimal for (3.3).

Before stating the result to be proved, we should highlight that under the assumption (H2), the stopping time τ_m is $\mathbb{P}_0$ and $\mathbb{P}_\infty$-a.s. finite. To see this, we should borrow some known results in the ruin theory. First, it is worth pointing out that the process $V(x)$ has a similar path of a dual risk process of an insurance company with an initial capital x. The latter is coupled with a strategy that pays out dividends to shareholders whenever the surplus level is beyond a certain barrier, which is known as the horizontal dividend barrier strategy, see Buhlmann (1970). Unlike the classical risk process which exhibits an upward drift with downward jumps due respectively to continuous positive premiums income and claims arrival, the dual process has an asymmetric path. It follows that the $m - V(x)$ has the same path as the dual formulation of the risk process. Making analogies with ruin theory, the barrier in our case is 0 and the ruin of the company occurs at time τ_m. If we consider the probability $\mathbb{P}_0(\tau_m < \infty)$ (resp. $\mathbb{P}_\infty(\tau_m < \infty)$), we can see that this coincides with the so-called ruin probability in the setting introduced above. Under $\mathbb{P}_0$, let $q_0(y)$ denote the probability of hitting the barrier m by the process $m - V(x)$, starting at y, before ruin, i.e. hitting the second barrier 0. It is shown in Segerdahl (1970) that this probability is given by

$$q_0(y) = \frac{1 - \psi_0(y)}{1 - \psi_0(m)},$$

where $\psi(y)$ is the ultimate ruin probability for the classical ruin problem, i.e. $\psi(y) = \mathbb{P}_0^{(m-y)}(\bar{\tau} < \infty)$, with $\bar{\tau} = \inf\{t \geq 0 : \; U_t > m\}$. In that case it is straightforward to show that τ_m, which the ruin time in presence of dividends is a.s. finite, we refer for example to Asmussen (2003).

Lemma 3. *The delay functional at τ_m is given by*

$$\mathbb{E}_\theta\left[\mathbf{1}_{\{\tau_m \geq \theta\}}(N_{\tau_m} - N_\theta)\Big|\mathcal{F}_\theta\right] = \mathbb{E}_\theta\left[\mathbf{1}_{\{\tau_m \geq \theta\}}\big(h_0(V_{\tau_m}) - h_0(V_\theta)\big)\big|\mathcal{F}_\theta\right]$$

Proof. First observe that Itô's lemma applied to h_0, the latter being continuously differentiable (Theorem 1), yields

$$h_0(V_t) = h_0(0) + \int_0^t (1-\rho)\lambda_s h_0'(V_{s-})ds + \int_0^t \big(h_0(V_s) - h_0(V_{s^-})\big)dN_s,$$

where $V(s^-)$ is the left-limit of V at s being equal to $V_s - \log\rho$ as seen earlier. Replacing this in the above equation gives

$$h_0(V_t) = h_0(0) + \int_0^t (1-\rho)\lambda_s h_0'(V_{s-})ds + \int_0^t \big(h_0(V_{s-} + \log\rho) - h_0(V_{s-})\big)dN_s. \tag{3.11}$$

Now, for any integer n let τ^n denote the stopping time defined as $\tau^n = \{t \geq \theta : (N_t - N_\theta) \geq n\}$. Replacing t by the stopping time τ_m^n at first and then by θ in (3.11), substituting one into the other and taking expectation with respect to $\mathbb{P}_\theta$

conditionally on $\mathcal{F}_\theta$ gives rise to

$$\mathbb{E}_\theta\Big[\mathbf{1}_{\{\tau_m^n\geq\theta\}}\big(h_0(V_{\tau_m^n})-h_0(V_\theta)\big)\big|\mathcal{F}_\theta\Big]=\mathbb{E}_\theta\Bigg[\mathbf{1}_{\{\tau_m^n\geq\theta\}}\int_\theta^{\tau_m^n}(1-\rho)\lambda_s h_0'(V_{s-})ds$$
$$+\mathbf{1}_{\{\tau_m^n\geq\theta\}}\int_\theta^{\tau_m^n}\big(h_0(V_{s-}+\log\rho)-h_0(V_{s-})\big)dN_s\Big|\mathcal{F}_\theta\Bigg].$$

On the other hand, note that on the subset $\{s\leq\tau_m^n\}$ we have $V_s\leq m$, so that Theorem 1 implies

$$h_0(V_s)-h_0(V_s+\log\rho)=\frac{1-\rho}{\rho}\big(h_0'(V_s)+1\big).$$

This, if substituted into the previous expression allows us to write

$$\mathbb{E}_\theta\Big[\mathbf{1}_{\{\tau_m^n\geq\theta\}}\big(h_0(V_{\tau_m^n})-h_0(V_\theta)\big)\big|\mathcal{F}_\theta\Big]=\mathbb{E}_\theta\Bigg[\mathbf{1}_{\{\tau_m^n\geq\theta\}}\int_\theta^{\tau_m^n}\frac{1-\rho}{\rho}h_0'(V_s)dM_s^\infty$$
$$+\mathbf{1}_{\{\tau_m^n\geq\theta\}}(N_{\tau_m^n}-N_\theta)\Big|\mathcal{F}_\theta\Bigg],$$

where M^∞ is the compensated $\mathbb{P}_0^{(x)}$-martingale, which is also a $\mathbb{P}_\theta^{(x)}$-martingale as soon as $\mathbb{P}_0^{(x)}$ and $\mathbb{P}_\theta^{(x)}$ coincides in the set $\{t\geq\theta\}$. Since h' is bounded, thanks to Lemma 2, the stochastic integral in the above expression vanishes (see Brémaud (1981)). Letting $n\to\infty$ in the last expression yields

$$\mathbb{E}_\theta\Big[\mathbf{1}_{\{\tau_m\geq\theta\}}\big(h_0(V_\theta)-h_0(V_{\tau_m})\big)\big|\mathcal{F}_\theta\Big]=\mathbb{E}_\theta\Big[\mathbf{1}_{\{\tau_m\geq\theta\}}(N_{\tau_m}-N_\theta)\Big|\mathcal{F}_\theta\Big],$$

using monotone convergence on the right-hand side and bounded convergence on the left. Noticing that $h_0(V_{\tau_m})=0$, the last equality can be rewritten as follows

$$\mathbb{E}_\theta\Big[\mathbf{1}_{\{\tau_m\geq\theta\}}(N_{\tau_m}-N_\theta)\Big|\mathcal{F}_\theta\Big]=\mathbb{E}_\theta\Big[\mathbf{1}_{\{\tau_m\geq\theta\}}h_0(V_\theta)\big|\mathcal{F}_\theta\Big]. \tag{3.12}$$

This concludes the proof. ☐

Remark 4. *Lemma 3 gains in interest if we realize that the event* $\{\tau_m\geq 0\}$ *may be written as* $\{V_s<m,\ s\leq\theta\}$*. Secondly, by the Markov property of the process* V*, from* (3.12)*, we have on the set* $\{\tau_m\geq\theta\}\cup\{V_s\}$

$$\mathbb{E}_\theta\Big[\mathbf{1}_{\{\tau_m\geq\theta\}}(N_{\tau_m}-N_\theta)\Big|\mathcal{F}_\theta\Big]=h_0(x),$$

which just amounts to saying that the conditional expectation in (3.12) *is the function only of the value* x *of* V_θ*. Moreover, as* h_0 *is decreasing for* $x\in[0,m]$ *we have* $\max_{0\leq x\leq m}h_0(x)=h_0(0)$ *which allows us to write*

$$\sup_{\theta\in[0,\infty]}\Big(\operatorname{ess\,sup}\mathbb{E}_\theta\Big[\mathbf{1}_{\{\tau_m\geq\theta\}}(N_{\tau_m}-N_\theta)\Big|\mathcal{F}_\theta\Big]\Big)=h_0(0). \tag{3.13}$$

This means loosely that the stopping time τ_m *is an equalizer rule over* θ*, in the sense that its performance does not depend on the value of the change-point. An other way of showing* (3.13) *is to see that the stopping time* τ_m *is a function of the process* V *only. Thus,* V *being Markov and due to* (2.8)*, we may conclude that* *all contribution of the process N_t before time θ may be summarized in V_θ. This together with Lemma 1 suggest that the worst case time to event delay before θ occurs whenever $V_\theta = 0$. More formally, we have*

$$\mathbb{E}_\theta\left[\mathbf{1}_{\{\tau_m \geq \theta\}}(N_{\tau_m} - N_\theta)\Big|\mathcal{F}_\theta\right] = \mathbb{E}_\theta\left[\mathbf{1}_{\{\tau_m \geq \theta\}}(N_{\tau_m} - N_\theta)\Big|V_\theta = 0\right] = \mathbb{E}_0\left[N_{\tau_m}\right]$$

.

This last remark shows that $R(\tau_m) = h_0(0)$, thus to prove the optimality of the stopping rule τ_m we will need to show that any stopping time satisfying the constraint in (3.4) has a worst case performance at least equal to $h_0(0)$. This is the purpose of what follows. However, before stating this we will need the following theorem. The latter is in the same vein as in Shiryaev (1996) and Moustakides (2004) and provides a convenient lower bound for the conditional worst case performance.

Theorem 2. *For any $\mathbb{P}_\infty$-a.s. finite stopping time τ, we have*

$$R(\tau) \geq \frac{\mathbb{E}_\infty\left[\int_0^\tau \exp\left(V_{s-}\right) dN_s\right]}{\mathbb{E}_\infty\left[\exp\left(V_\tau\right)\right]}. \tag{3.14}$$

Proof. Let τ be a $\mathbb{P}_\infty$-a.s. finite stopping time and $H = (H_t)_{t\geq 0}$ an $\mathbb{F}$-adapted process then we have

$$\begin{aligned} R(\tau)\mathbb{E}_\infty\left[H_\tau\right] &= \mathbb{E}_\infty\left[R(\tau)H_\tau\right] = \mathbb{E}_\infty\left[R(\tau)\int_0^\tau \mathbf{1}_{\{\theta\leq\tau\}}dH_\theta\right], \\ &\geq \mathbb{E}_\infty\left[\int_0^\infty R_\theta(\tau)\mathbf{1}_{\{\theta\leq\tau\}}dH_\theta\right], \end{aligned}$$

such that $R_\theta(\tau)$ is given by

$$\begin{aligned} R_\theta(\tau) &= \mathbb{E}_\theta\left[(N_\tau - N_\theta)\mathbf{1}_{\{\tau\geq\theta\}}\big|\mathcal{F}_\theta\right] \\ &= \int_0^\infty \mathbb{E}_\theta\left[\mathbf{1}_{\{s\leq\tau\}}dN_s\big|\mathcal{F}_\theta\right] = \mathbb{E}_\infty\left[\int_0^\tau e^{U_{s-}-U_{\theta-}}dN_s\big|\mathcal{F}_\theta\right]. \end{aligned}$$

Substituting the above inequality becomes

$$\begin{aligned} R(\tau)\mathbb{E}_\infty\left[H_\tau\right] &\geq \mathbb{E}_\infty\left[\int_0^\infty \mathbf{1}_{\{\theta\leq\tau\}}\mathbb{E}_\infty\left[\int_\theta^\tau e^{U_{s-}-U_{\theta-}}dN_s\big|\mathcal{F}_\theta\right]dH_\theta\right] \\ &= \mathbb{E}_\infty\left[\int_0^\tau \mathbb{E}_\infty\left[\int_\theta^\tau e^{U_{s-}}dN_s\big|\mathcal{F}_\theta\right]d\tilde{H}_\theta\right], \\ &= \mathbb{E}_\infty\left[\int_0^\tau \mathbb{E}_\infty\left[\xi_\tau\big|\mathcal{F}_\theta\right]d\tilde{H}_\theta\right] - \mathbb{E}_\infty\left[\int_0^\tau \xi_\theta d\tilde{H}_\theta\right], \end{aligned}$$

such that $\tilde{H}_t = \int_0^t e^{-U_{s-}}dH_s$ and $\xi_t = \int_0^t e^{U_{s-}}dN_s$.
Let $Y_\theta = \mathbb{E}_\infty[\xi_\tau|\mathcal{F}_\theta]$, using the formula for integration by parts gives rise to

$$Y_t\tilde{H}_t - Y_0\tilde{H}_t = \int_0^t Y_{\theta-}d\tilde{H}_\theta + \int_0^t \tilde{H}_{\theta-}dY_\theta + \sum_{0\leq\theta\leq t}\Delta Y_s\Delta\tilde{H}_s.$$

Inserting

$$\sum_{0\leq\theta\leq t}\Delta Y_s\Delta\tilde{H}_s = \int_0^t \Delta Y_\theta d\tilde{H}_\theta = \int_0^t (Y_\theta - Y_{\theta-})d\tilde{H}_\theta,$$

in the last expression yields

$$\int_0^t Y_\theta d\tilde{H}_\theta = Y_t \tilde{H}_t - \int_0^t \tilde{H}_{\theta-} dY_\theta. \tag{3.15}$$

Since the process Y is a non-negative $\mathbb{P}_\infty$-martingale, then by the general theory of stochastic integration, $\int_0^t \tilde{H}_\theta dY_\theta$ is a $\mathbb{P}_\infty$-local martingale (see (Brémaud 1981, Theorem 6, p. 10)). Then letting $t = \tau$ in (3.15) and taking expectation with respect to $\mathbb{P}_\infty$, it follows

$$\mathbb{E}_\infty \left[\int_0^\tau Y_\theta d\tilde{H}_\theta \right] = \mathbb{E}_\infty \left[Y_\tau \tilde{H}_\tau \right],$$

where we used the optional stopping theorem in the left-hand side of (3.15). More precisely, we have $\mathbb{E}_\infty \left[\int_0^\tau \tilde{H}_\theta dY_\theta \right] = 0$. Substituting the latter into the previous inequality allows us to write

$$\begin{aligned} R(\tau)\mathbb{E}_\infty [H_\tau] &\geq \mathbb{E}_\infty \left[\xi_\tau \tilde{H}_\tau \right] - \mathbb{E}_\infty \left[\int_0^\tau \xi_\theta d\tilde{H}_\theta \right] \\ &= \mathbb{E}_\infty \left[\int_0^\tau \tilde{H}_\theta d\xi_\theta \right] = \mathbb{E}_\infty \left[\int_0^\tau \int_0^\theta e^{-U_{s-}+U_{\theta-}} dH_s dN_\theta \right]. \end{aligned} \tag{3.16}$$

Let H be defined as follows

$$H_t = e^{V_t} - 1 - (1-\rho) \int_0^t e^{V_{s-}} dM_s^\infty, \tag{3.17}$$

for every $t \geq 0$. Optional stopping at time τ yields

$$\mathbb{E}_\infty[H_\tau] = \mathbb{E}_\infty \left[e^{V_\tau} \right] - 1 - \mathbb{E}_\infty \left[\int_0^\tau e^{V_{s-}} dM_s^\infty \right] = \mathbb{E}_\infty \left[e^{V_\tau} \right] - 1. \tag{3.18}$$

Now, using (2.8) and applying Itô's rule to (3.17) we can easily check that

$$e^{-U_{s-}} dH_s = -(1-\rho) e^{-U_{s-}+V_{s-}} \mathbf{1}_{\{V_{s-}=0\}} \lambda_s ds = -e^{L_s} \mathbf{1}_{\{V_{s-}=0\}} \left(dU_s - \log \rho dN_s \right).$$

Integrating the latter gives

$$\int_0^t e^{-U_{s-}} dH_s = -\int_0^t e^{L_s} \mathbf{1}_{\{V_{s-}=0\}} dU_t + \int_0^t e^{L_s} \mathbf{1}_{\{V_{s-}=0\}} \log \rho dN_s.$$

Note that L_s only increases when $V_s = 0$, in which case $L_s = -U_s$, so that the first stochastic integral in the right-hand side reduces to $e^{L_t} - 1$. The second integral vanishes as far as N_s is constant when $V_s = 0$ for every $s \geq 0$, so we can write

$$\int_0^t e^{-U_{s-}+U_t} dH_s = e^{U_t + L_t} - e^{U_t} = e^{V_t} - e^{U_t}. \tag{3.19}$$

Substituting (3.18) and (3.19) into (3.16), we obtain

$$R(\tau)(\mathbb{E}_\infty \left[e^{V_\tau} \right] - 1) \geq \mathbb{E}_\infty \left[\int_0^\tau \left(e^{V_s} - e^{U_{s-}} \right) dN_s \right]. \tag{3.20}$$

Now, note that for any bounded stopping time τ we have

$$R(\tau) = \sup_\theta \left(\operatorname{ess\,sup} \mathbb{E}_\theta \left[(N_\tau - N_\theta)^+ \middle| \mathcal{F}_\theta \right] \right) \geq \mathbb{E}_0 \left[N_\tau \right] = \mathbb{E}_\infty \left[e^{U_\tau} N_\tau \right].$$

Similarly to (3.15), it is easy to check that $\mathbb{E}_\infty\big[e^{U_\tau N_\tau}\big] = \mathbb{E}_\infty\big[\int_0^\tau e^{U_{t-}} dN_t\big]$, and thus

$$R(\tau) \geq \mathbb{E}_\infty\Big[\int_0^\tau e^{U_{t-}} dN_t\Big]. \tag{3.21}$$

Addition of (3.20) and (3.21) provides the desired inequality. □

We are now in a position to state the optimality of the stopping rule τ_m for the problem in (3.3).

Theorem 3. *Any stopping time τ satisfying the false alarm constraint in (3.4) such that $\mathbb{E}_\tau[N_\tau] \geq \mathbb{E}_\tau[N_{\tau_m}]$ has a worst case time to event that is no less than $h_0(0)$, i.e.*

$$R(\tau) \geq h_0(0). \tag{3.22}$$

Before proceeding to the proof of the theorem, the following additional lemma will be needed.

Lemma 4. *For $0 \leq x \leq m$, we have*

$$\mathbb{E}_\infty\left[\rho h_\infty(V_\tau) - e^{V_\tau} h_0(V_\tau)\right] \geq 0.$$

Proof. Define the function $g(x) = \rho h_\infty(x) - e^x h_0(x)$. The latter being continuously differentiable and uniformly bounded, see Theorem 1 and Lemma 2, Itô's rule applies and we obtain

$$\begin{aligned} dg(V_t) &= (1-\rho)\lambda_t g'(V_{t-})\mathbf{1}_{\{V_{t-}>0\}}dt + (g(V_{t-}+\log\rho) - g(V_{t-}))dN_t \\ &= (1-\rho)\lambda_t g'(V_{t-})\mathbf{1}_{\{V_{t-}=0\}}dt + (1-\rho)\lambda_t g'(V_{t-})dt \\ &\qquad + \big(g(V_{t-}+\log\rho) - g(V_{t-})\big)dN_t \\ &= (1-\rho)\lambda_t g'(V_{t-})\mathbf{1}_{\{V_{t-}=0\}}dt - (1-\rho)g'(V_{t-})dM_t^\infty \\ &\qquad + \big((1-\rho)g'(V_{t-}) + g(V_{t-}+\log\rho) - g(V_{t-})\big)dN_t \end{aligned}$$

Integrating up to τ and taking expectation with respect to $\mathbb{P}_\infty$ yields

$$\mathbb{E}_\infty\big[g(V_\tau)\big] = \mathbb{E}_\infty\Big[\int_0^\tau \big(e^{V_t}-1\big)dN_t\Big] + g(0).$$

The first term in the left-hand side is positive as far as $V_t \geq 0$ for $t \geq 0$. For the second term $g(0)$, notice that

$$\begin{aligned} \rho\mathbb{E}_\infty\big[N_{\tau_m^n(x)}\big] = \mathbb{E}_\infty\big[e^{U_{\tau_m^n(x)}}N_{\tau_m^n(x)}\big] &= \mathbb{E}_\infty\big[\rho\Lambda_{\tau_m^n(x)}\big] \\ &= \mathbb{E}_0\big[\rho e^{U_{\tau_m^n(x)}}\Lambda_{\tau_m^n(x)}\big] \\ &\geq \mathbb{E}_0\big[\rho e^{U_{\tau_m^n(x)}}\Lambda_{\tau_m^n(x)}\big] \\ &\geq \mathbb{E}_0\big[\rho\Lambda_{\tau_m^n(x)}\big] = \mathbb{E}_0\big[N_{\tau_m^n(x)}\big]. \end{aligned}$$

In the last inequality we used the fact that $V_{\tau_m(x)} = U_{\tau_m(x)} > 0$ and thus $U_{\tau_m^n(x)} \geq 0$ $\mathbb{P}_0$-a.s. respectively. Finally we have $g(0) \geq 0$ which leads to the required inequality. □

Proof. Based on Theorem 2 it suffices to show that

$$R(\tau) \geq \frac{\mathbb{E}_\infty\left[\int_0^\tau e^{V_s} dN_s\right]}{\mathbb{E}_\infty\left[e^{V_\tau}\right]} \geq R(\tau_m) = h_0(0),$$

where the last equality is owed to Remark 4. This becomes

$$\mathbb{E}_\infty\left[\int_0^\tau e^{V_s} dN_s - e^{V_\tau} h_0(0)\right] \geq 0.$$

From the particular property of the path of V we have $e^{V_s} = \rho e^{V_{s-}}$, which gives, when substituted in (3.23)

$$\mathbb{E}_\infty\left[\int_0^\tau \rho e^{V_s} dN_s - e^{V_\tau} h_0(0)\right] \geq 0. \tag{3.23}$$

Define the function f, each $0 \leq x < m$, as $f(x) = \rho(h_\infty(x) - h_\infty(0)) - e^x(h_0(x) - h_0(0))$. Notice that f is continuously differentiable and uniformly bounded, see Theorem 1 and Lemma 2. Thanks to (3.7) and (3.8) we can easily check that

$$(1 - \log\rho) f'(x) + f(x + \log\rho) - f(x) = \rho\big(e^x - 1\big). \tag{3.24}$$

Applying Itô's rule to $f(V_t)$ yields

$$\begin{aligned} df(V_t) &= (1-\rho)\lambda_t f'(V_{t-}) \mathbf{1}_{\{V_{t-}>0\}} dt + \big(f(V_{t-} + \log\rho) - f(V_{t-})\big) dN_t \\ &= (1-\rho)\lambda_t f'(V_{t-}) \mathbf{1}_{\{V_{t-}=0\}} dt - (1-\rho) f'(V_{t-}) dM_t^\infty \\ &\qquad + \Big(f(V_{t-} + \log\rho) - f(V_{t-}) + (1-\rho) f'(V_{t-})\Big) dN_t. \end{aligned}$$

Substituting (3.24) in the above expression, integrating up to τ and taking expectation with respect to $\mathbb{P}_\infty$ yields

$$\begin{aligned} \mathbb{E}_\infty\big[f(V_\tau)\big] &= \mathbb{E}_\infty\Big[(1-\rho)\int_0^\tau \lambda_t f'(V_{t-}) \mathbf{1}_{\{V_{t-}=0\}} dt - (1-\rho)\int_0^\tau f'(V_{t-}) dM_t^\infty \\ &\qquad + \int_0^\tau \big(e^{V_{t-}} - 1\big) dN_t\Big]. \end{aligned}$$

where we used the fact that $f(0) = 0$. Analysis similar to that in the proof of Theorem 2 shows that the first two integrals vanishes. We thus get

$$\begin{aligned} \mathbb{E}_\infty\Big[\int_0^\tau \big(e^{V_{t-}} - 1\big) dN_t\Big] &= \mathbb{E}_\infty\Big[h_\infty(V_\tau) - h_\infty(0) - e^{V_\tau}\big(h_0(V_\tau) - h_0(0)\big)\Big] \\ &= \mathbb{E}_\infty\big[f(V_\tau)\big] \end{aligned} \tag{3.25}$$

Having disposed of this preliminary step, we can now return to proof of the theorem. Inserting (3.23) into the left-hand side of (3.25) gives rise to

$$\begin{aligned} \mathbb{E}_\infty\Big[\int_0^\tau \rho e^{V_{s-}} dN_s - e^{V_\tau} h_0(0)\Big] &= \mathbb{E}_\infty\Big[f(V_\tau) - e^{V_\tau} h_0(0)\Big] \\ &= \mathbb{E}_\infty\Big[\rho\big(h_\infty(V_\tau) - h_\infty(0)\big) - e^{V_\tau}\big(h_0(V_\tau) - h_0(0)\big) \\ &\qquad - e^{V_\tau} h_0(0) + \rho N_\tau\Big]. \end{aligned}$$

Rearranging terms leads to

$$\mathbb{E}_\infty\Big[\int_0^\tau \rho e^{V_{s-}} dN_s - e^{V_\tau} h_0(0)\Big] = \rho\Big(\mathbb{E}_\infty\big[N_\tau\big] - h_\infty(0)\Big) \qquad (3.26)$$
$$+\Big(\mathbb{E}_\infty\big[\rho h_\infty(V_\tau) - e^{V_\tau} h_0(V_\tau)\big]\Big).$$

Note the first term in the right-hand side of (3.26) is positive, as $\mathbb{E}_\infty[N_\tau] \geq h_\infty(0)$, as well as the second term which is due to Lemma 4. This shows (3.23) and concludes the proof. □

4. Numerical Analysis

In this section, we investigate the efficiency of the detection procedure. First, we illustrate some properties of the procedure by means of Monte-Carlo simulation, using different proportional alternatives. Next, we will be interested in insurance-related applications, and more precisely the detection in the detection of structural breaks in the proportional hazard rate two-population model previously introduced. It is worth recalling that the CUSUM detection procedure can be split into the following steps:

Step 1: Fix the input parameters: The post-change intensity through the specification of ρ and the false alarm constraint ω.

Step 2: Determine the threshold m as the solution of the equation $\mathbb{E}_\infty[N_{\tau_m}] = \omega$. This can be done whether by using the methodology in Section 5 or by means of Monte-Carlo simulations.

Step 3: For each new observation at time t compute the value of the CUSUM process V given by the iterative relation $V_{t+1} = (V_{t-1} + U_t)^+$.

Step 4: Compare the current value of V to the threshold m and stop the procedure once $V_t \geq m$ and sound an alarm. Hence $\tau_m = t$.

4.1. **Simulation Study.**

We apply the detection scheme to monitor breaks arising from a proportional relationship between two intensities. The baseline time-dependent intensity denoted by λ_t has a break-point which we fix for the purpose of our simulation at $\theta = 0$. After the change, the new intensity is given by $\rho\lambda_t$. As inputs of the procedure, we need to characterize the process through the intensity λ and the proportional parameter ρ after the change-point. Recall that in order to follow the scheme given above, one should consider an increasing intensity in order to satisfy Assumption (H2). By increasing intensity we mean that the population under study is growing over time. This is given as follows:

$$\lambda_t = l_0 \exp(b(c^t - 1)) + B_t, \qquad (4.1)$$

where l_0 is the initial size of the population, $\exp(b(c^t - 1))$ is Gompertz curve of mortality that we need to fit to the data (b and c being constant). Finally, B_t gives the arrival of individuals with a constant rate δ.

Clearly, the performance of the detection will depend on our faculty to properly identify the change-point, which depends on the size of breakpoint namely ρ, l_0 and

δ and of course on the Gompertz curve $\exp(bc^t)$. In the following, we only report results for different assumptions on ρ and l_0. By doing so we only intend to give key parameters that can be monitored by an insurance, the remaining parameters are either exogenous or imposed such as the mortality evolution. For this reason, we decide to target shift sizes $\rho = 1.10$ and 1.25, corresponding respectively to 10% and 25% increases in the intensity. Furthermore, the initial population size is set to $l_0 = 100, 10000$ and 100000. The arrival rate $\delta = 0.8$ and $b = 0.001$ and $c = 1.01$ are fixed.

On the other hand, as far as the false alarm constraint is concerned,one may face a dilemma to properly implement the procedure. Although it is more meaningful for an insurer to minimize the excess of mortality after the occurrence of a change, which makes sense from liability management point of view, one may not be able to guess directly an appropriate threshold for the false alarm constraint described in terms of the number of events until sounding a false alarm. Thus we shall rather use the traditional time being more meaningful in terms of false alarm constraint of the procedure, and follow the same idea in Gandy et al. (2010). To do so, it suffices to recall the so-called Wald's identity given by $\mathbb{E}_\infty[N_{\tau_m}] \approx \tilde{\lambda}\mathbb{E}_\infty[\tau_m]$, where $\tilde{\lambda} = \lim_{t\to\infty} \mathbb{E}_\infty[N_t]/t$, see also Moustakides (1999). Next, we will fix the parameter $\tilde{\lambda}$ as the mean intensity over the calibration period. Table 1 reports

l_0	100		10000		100000	
ω	100	500	100	500	100	500
$rho = 1.10$	431	593	129	192	44	102
$\rho = 1.25$	389	435	98	177	27	89

Table 1: *Monte-Carlo simulation of the average events (deaths) until detection*

the average detection delay for different parameters sets, with 20000 replications. First, we notice that the detection delay decreases with the size of the shift on the intensity, which is a typical behaviour of the CUSUM rule and also very common for most detection rules. We also note that the detection delay increases with the false alarm constraint, as one needs more observation on deaths to detect an initial shift. This is due the fact under small false alarm constraints ω we are more likely to declare false changes so increasing ω will postpone the detection. On the other hand, there are some disparities between the three assumptions on the initial size of the population. Indeed, we notice that the detection delay decreases with the size the population. Thus the number of observed deaths needed to declare the change is highly dependent on the population size. To see this, recall that sounding the alarm is made when the reflected process V hits the barrier m. The latter is attaining m by means of jumps, so as far as the population size is small the deaths are less frequent and V takes more time to attain the barrier. Thus one should keep in mind that the detection using the time until event criterion is highly dependent on the population under study. We can also expect that for high mortality profile population the detection time for changes may be longer. In the following subsection, we explore the detection of changes on a real database.

4.2. **Real-World Application.** We consider a real-world pension fund policyholder portfolio with an estimated mortality profile given by the historical death records. In order to assess the detected changes we need some backing results for comparison. Indeed, it is hard to decide if the detected change corresponds to a *real* change or to a false alarm. For this reason, we first implement an off-line procedure to locate the change-points in the data. Then, we perform the online detection and we compare results of the two approaches.

We consider the whole population of England & Wales and the assured live mortality. We consider a training period used to estimate (2.1). Hence, we set-up a surveillance period accounting for population growth, i.e. new entries in the portfolio. The latter coincides with entrance of the surviving individuals in the cohort neighboring our age-groups. As insurer are used to work on 10-year age group we consider the following age groups: $\{50-59, 60-69, 70-79, 80-89\}$.

4.2.1. *Off-line change-points detection.* The off-line detection focuses on the detection of the changes in the local characteristics and the estimation of the location of change-points (analogous to θ in our approach). This is generally done thanks to *a posteriori* hypothesis testing. More formally, we suppose that the process λ_t/λ_t^0 is observed over a period $[0, T]$, and the problem of interest is the detection of changes[2] in the local characteristics of ρ, and the estimation of the times when changes occur. This off-line hypothesis testing problem traditionally uses a trade-off between the ability to detect actual changes when they occur, called the *power* of procedure, and the ability not to detect anything when no change occurs, which is related to the type I error. Many frameworks were investigated in the literature of the off-line detection. For example, the so-called *Filtered derivative* is based on a moving average estimator of ρ, see Fhima et al. (2011) and Bertrand et al. (2011). Wavelet based estimation of the trend for localized signal detection is explained in Siegmund and Worsley (1995). Fhima et al. (2011) and Bertrand et al. (2011) investigate the changes over a fixed period using a linear estimator of the latter on two sliding windows respectively at the right and at the left for every location. Consequently, the test statistic is chosen as the maximum of the difference between the two estimation. However, and unlike Siegmund and Worsley (1995), the filtering procedure is operated using a pre-defined window of size which stems from expert opinions. More formally, if we denote $X = (X_t)_{t\geq 0}$ the ratio of the hazard rates $X_t = \lambda_t/\lambda_t^0$, the following statistic is used to locate the change

[2]Note that for off-line detection we are looking for one or more change-points.

point

$$X_{\max} = \max_{\theta,s} s^{-1/2} \int f\big(s^{-1}(t-\theta)\big) dX_t, \tag{4.2}$$

where f is wavelet-like function and s is the scale parameter. The integral is computed over the observation period. Thus, the wavelet transform in (4.2) quantifies a difference of adjacent averages in the neighborhood of each location t at scale s (which represents the size of the estimation window). It is shown that this statistic corresponds to the maximum log-likelihood ratio statistic for the following hypothesis testing problem:

$$dX_t = \rho + ms^{-1/2} f\big((t-t_0)/s\big)dt + dB_t, \tag{4.3}$$

where B_t is a standard Brownian motion. In that case we are testing for $m \neq 0$ and f stands for the shape of the irregularity in the trend ρ. A special case of (4.2) is when f is the Haar wavelet. We come up with a similar statistic as in Bertrand et al. (2011) but for a set of windows. As we do not know the shape of the transition between consecutive trends, we use the Gaussian wavelet given by

$$f(t) = \pi^{-1/4} \exp\big(-t^2/2\big), \tag{4.4}$$

where this not only accounts for abrupt changes but may also allow us to detect gradual changes. The test is now constructed to detect changes on the smoothed process by searching over the filter width s as well as the time location θ's. In Worsley (2001), potential change-points are selected as corresponding to the times θ and scales s where the value of the test statistic $X_{\max}$ exceeds a given threshold h. This is equivalent to look for regions in the time-scale space where the null hypothesis of *no-change* is rejected. The test leads to formulate the condition of the rejection of the absence of change, $H_0 : \ m = 0$. This yields to reject the hypothesis for regions of the time-scale space when the probability of type I error exceeds a critical level p^*:

$$\mathbb{P}\left(X_{\max} > \theta \mid H_0 \text{ is true }\right) = p^*, \tag{4.5}$$

such that the threshold h is the critical value. We should then fix a probability of type I error at level p^*, and determine the critical value h from (4.5), which assesses the validity of the null hypothesis and also permits to identify the region where the trend change is present. Usually, there is no closed form formula for the probability in (4.5) as we only have the asymptotic distribution of the test statistic $X_{\max}$. We use an approximation of (4.5) proposed in Worsley (2001). This approximation uses the topological properties of excursion regions defined as $X_{\max} \geq h$. This is done using the so-called expected *Euler Characteristic* which approaches the expected number of local maxima as the threshold h increases.
In Figure 2, we depict the wavelet-like transform in (4.2). The solid curves delimit the regions of time-scale space (t, s) where the hypothesis of no change in ρ is rejected. The critical values of the test are determined using Worsley (2001). This figure shows that many changes are present in the process X_t. First, we may

separate the scale axis into two regions. Changes at small scales correspond to local irregularities (peaks, jumps, etc.). Changes at high scales are more likely to be changes in the long term as the estimation is done over a longer window. We notice that for high scales, a change is detected. The most likely change instant is identified as the time where $X_{\max}$ is attained, following the approach of Worsley (2001). In our case, this corresponds to 1971, 1977, 1974 and 1973 respectively for age groups $50-59$, $60-69$, $70-79$ and $80-89$.

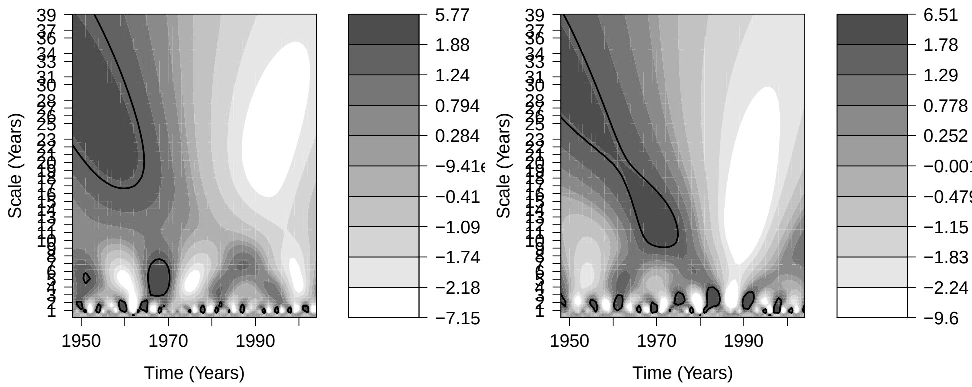

Figure 2: *Wavelet transform of the process λ_t/λ_t^0 with a Gaussian-wavelet. The solid curve shows the regions where the null hypothesis of no change in the parameter ρ is rejected, for two age groups: 50-59 (left) and 80-89 (right)*

4.2.2. *Cusum detection and comparison.* We are given a portfolio insured population which comes from the Continuous Mortality Investigation (CMI) assured population derived from data submitted by U.K. life insurers. We aim at linking this population to the whole national England & Wales population through a proportional hazard model. Next, to implement the surveillance schemes in the examples below, we need to specify the post-change coefficient ρ. This is not an easy task, especially in the context of life insurance risk management. To have some estimated values, which are not necessarily the best ones, for these parameters, one possible approach is to use a training period $1947-1969$ for instance, and then to estimate the pre-change rate relationship and the post-change coefficient is set respectively to 15% and 50%. These choices are intended only for illustrations.
For the each age group, we start the CUSUM schemes immediately after the period we used for fitting and run them until hitting the threshold level m. The latter was determined by equating the false alarm constraint with $\omega = 100\bar{\lambda}$, $\bar{\lambda}$ being the average intensity over the fitting period.

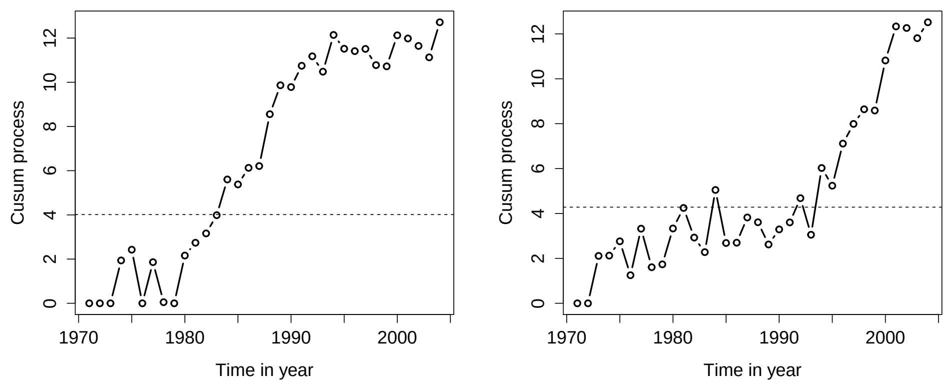

Figure 3: *Detection scheme for age groups* $50-59$ *(right) and* $80-89$ *(left). The post-change is set to* $\rho = 15\%$ *and the false alarm constraint to* $\omega = 100\bar{\lambda}$*.*

	τ_m		
Age	$\rho = 1.50$	$\rho = 1.15$	Observed
$50-59$	1984	1978	1971
$60-69$	1991	1985	1974
$70-79$	1988	1984	1974
$80-89$	1983	1978	1973

Table 2: *Detection of mortality change with a post-change ratio of* $\rho = 1.15$ *and an average run length (false alarm) constraint of* 100*. The right column reports the detected change-point using an off-line procedure.*

In Table 2 we report the detected change-point, see also Figure 3. All schemes triggered an alarm during the surveillance period for both assumptions on ρ. These change-points may correspond to those already located using the off-line procedure. Note that when ρ is larger, the detection delay is logically longer.
The detected change-points are implicitly mentioned in Coughlan et al. (2010). After working on the same dataset, they notice that the assured lives data has been changing since 1980. It is worth mentioning that this data was collected from E. & W. insurers, and that around the 80's some insurers stopped providing their own information. This fact and the results of the off-line detection make us confident that the changes detected online are unlikely to correspond to false alarms.
Note that in the present examples we only have yearly observations of the death counting process of the assured lives portfolio. We know N_t and N_{t+1} for each

integer t in the observation period, but we do not observe N_s for non-integer instants s. For the sake of simplicity, using ruin theory vocabulary, we have simplified the dividend type mechanism: we assume that dividends are paid at the end of each period. This may slightly alter the behavior of the CUSUM process, mainly when the process is close to reflection barrier. However it is possible to show that this modification has an impact of magnitude smaller than $\log \rho$ on the process, if we replace (uniformly distributed when λ is constant) conditional jump instants with deterministic ones (at times $t + k/(n+1)$, $1 \leq k \leq n$ if there are n jumps in $[t, t+1)$). Consequently, this approximation can slightly delay detection in some particular cases. So, if one has all the information and is able to reconstruct the entire sample path, one might even detect changes slightly faster. This will be the case for pension funds and insurance companies, who observe exact death instants, sometimes after one month or two. For applications to Swiss Re type products, one is interested in differences between two population indices that are only available every year or every term. In that case, one must either use an approximation like the one we propose, or work with a (theoretical) probability of detection given discrete observations using excursion theory.

5. Conclusion

In this paper, we have shown the optimality of the CUSUM stopping, to detect abrupt changes in the Cox (1972) proportional hazard two-population model, for a Lorden-like delay criterion. Besides, we have been able to show the continuous differentiability of the performance measures as well as partial differential equations fulfilled by these functions. We also investigate the performance of detection using some parameters sets. The detection is highly dependent on the initial size of the population given an underlying mortality dynamics.

Throughout the paper, we have assumed that the intensity is deterministic. One could adopt the same arguments used in this paper to show the optimality in the case of stochastic intensity. Indeed, we should consider an augmented filtration generated both by the death counts and the observations of the intensity. A typical example is an intensity governed by a standard Brownian motion $W = (W_t)_{t \geq 0}$. The augmented filtration can be specified for each $t \geq 0$ as follows:

$$\mathcal{H}_t = \mathcal{F}_t \vee \mathcal{G}_t,$$

where $G_t = \sigma\{s \leq t, W_s\}$. Given that $\mathbb{F}$ and $\mathbb{G}$ are independent, we come up with the same type of log-likelihood ratio as before, and we could extend our detection tool to more sophisticated cases. This is left for further research.

Appendix

Expected events until detection

In order to compute the number of events until sounding an alarm, respectively under $\mathbb{P}_0$ and $\mathbb{P}_\infty$, we will focus on the new processes $\tilde{N}$ and $\tilde{V}(x)$. The latter being given under $\mathbb{P}_0$ by

$$\tilde{V}_t(x) = x + \underbrace{(\log(\rho)N_t + (1-\rho)t)}_{\tilde{U}_t} + \underbrace{(-x - \underline{U}_t)^+}_{\tilde{L}_t(x)}. \tag{5.1}$$

From Kella and Whitt (1992) (see also Asmussen (2003)), the following process

$$M_t^\alpha = \phi(\alpha)\int_0^t e^{\alpha \tilde{V}_s(x)} ds + e^{\alpha x} - e^{\alpha \tilde{V}_t(x)} + \alpha \int_0^t e^{\alpha \tilde{V}_s(x)} d\tilde{L}_s(x), \tag{5.2}$$

is a $\mathbb{P}_0$-martingale, where the counting process is a standard Poisson process with intensity 1. Here, the function ϕ is the levy exponent of the process $\tilde{N}$ given by $\phi(\alpha) = \mathbb{E}_\infty[e^{\alpha \tilde{N}_1}]$. As noted earlier, the process $\tilde{L}_t(x)$ only increases when $\tilde{V}_s(x) = 0$, which makes the integral in (5.2) reduces to $\alpha \tilde{L}_t(x)$. Since $\tilde{L}_0(x) = 0$, it follows that the process M^α is a zero mean martingale. Next, replacing t by $\tilde{\tau}_m(x)$ and applying the optional sampling in (5.2) while taking the expectation with respect to $\mathbb{P}_0$ yields

$$\mathbb{E}_0[\tilde{V}_{\tilde{\tau}_m(x)}(x)] = x + \phi'(0)\mathbb{E}_0[\tilde{\tau}_m(x)] + \mathbb{E}_0[L_{\tilde{\tau}_m(x)}]. \tag{5.3}$$

On the other hand, we recall from (5.1) that

$$\begin{aligned} \mathbb{E}_0[\tilde{V}_{\tilde{\tau}_m(x)}(x)] &= x + \log\rho\, \mathbb{E}_0[\tilde{N}_{\tilde{\tau}_m(x)}] + (1-\rho)\mathbb{E}_0[\tilde{\tau}_m(x)] \\ &= \Big(\log\rho + 1 - \rho\Big)\mathbb{E}_0[\tilde{\tau}_m(x)], \end{aligned}$$

where the last equality is due the fact that $\tilde{N}_t - \tilde{t}$ is a zero mean martingale. Substituting the right-hand side of the last equality in (5.3) and rearranging terms gives arise to

$$\mathbb{E}_0[\tilde{\tau}_m(x)] = \frac{\mathbb{E}_0[L_{\tilde{\tau}_m(x)}(x)]}{\log\rho + 1 - \rho + \phi'(0)}. \tag{5.4}$$

In order to derive the expected time until hitting the barrier m we should find $\mathbb{E}[L_{\tilde{\tau}_m(x)}(x)]$.

It is well known that, see Irbäck (2003) and Dickson and Gray (1984) that $\mathbb{E}_0[L_{\tilde{\tau}_m(x)}(x)] = q_{m-x}(1-\rho)/p$; q_{m-x} being the probability of hitting 0 before attaining the barrier m, and $p = 1 - q_0$. It is also shown that

$$q_{m-x} = \frac{1 - \psi(m-x)}{1 - \psi(0)},$$

where $\psi(m-x)$ is the ultimate ruin probability when the initial capital $m-x$. We are now left with the task of determining the ultimate ruin probability with constant size of claims. To this end, let us recall the so-called Beckman's convolution formula defined for $0 \geq x \leq m$ as

$$\psi(x) = 1 - \frac{\log\rho}{1-\rho}\sum_{n=0}^{\infty}\left(\frac{\log\rho}{1-\rho}\right)^n F_e^{*n}(x), \tag{5.5}$$

where $F_e^{*n}(x)$ is the n-fold convolution of $F_e(x) = \frac{1}{\log\rho}\int_0^x \mathbf{1}_{\{y\leq\log\rho\}}$. Let f be the uniform density function over $[0,\log\rho]$, then it follows that (5.5) may be written as

$$\psi(x) = 1 - \frac{\log\rho}{1-\rho}\sum_{n=0}^{\infty}\left(\frac{\log\rho}{1-\rho}\right)^n \int_0^x f^{*n}(y)dy. \tag{5.6}$$

The n-fold convolution of f is given can be computed as follows [see Renyi (1970)]

$$\begin{aligned} f^{*n}(x) &= \sum_{j=0}^{\lceil\frac{x}{\log\rho}\rceil}(-1)^j\frac{n}{j!(n-j)!}\left(\frac{x}{\rho}-j\right)^{n-1}, \\ &= \sum_{j=0}^{\lceil\frac{x}{\log\rho}\rceil}\frac{d}{dx}\left\{(-1)^j\frac{1}{j!(n-j)!}\left(\frac{x}{\rho}-j\right)^{n}\right\}, \end{aligned}$$

where $\lceil x/\log\rho\rceil$ is the largest integer less than or equal to $x/\log\rho$. Substituting the right-hand side of the last equation in (5.6) Interchanging the order of summation, we obtain more simply:

$$\begin{aligned} 1-\psi(x) &= \frac{\log\rho}{1-\rho}\sum_{j=0}^{\lceil\frac{x}{\log\rho}\rceil}\int_0^x\frac{d}{dy}\left\{\sum_{n=0}^{\infty}\left(\frac{\log\rho}{1-\rho}\right)^n\{\frac{1}{(n-j)!}\left(\frac{y}{\rho}-j\right)^n\right\} \\ &= \frac{\log\rho}{1-\rho}\sum_{j=0}^{\lceil\frac{x}{\log\rho}\rceil}\frac{(-1)^j}{j!}\int_0^x\frac{d}{dy}\left\{\exp\left(\frac{\log\rho}{1-\rho}\left(\frac{y}{\rho}-j\right)\right)\right\} \\ &= \frac{\log\rho}{1-\rho}\sum_{j=0}^{\lceil\frac{x}{\log\rho}\rceil}\frac{(-1)^j}{j!}\int_0^x\frac{d}{dy}\left\{\exp\left(\frac{\log\rho}{1-\rho}\left(\frac{y}{\rho}-j\right)\right)\right\} \\ &= \frac{\log\rho}{1-\rho}\sum_{j=0}^{\lceil\frac{x}{\log\rho}\rceil}\frac{(-1)^j}{j!}\left\{\exp\left(\frac{\log\rho}{1-\rho}\left(\frac{x}{\rho}-j\right)\right)\right\} \end{aligned}$$

References

Asmussen, S. 2003. *Applied probability and queues*, vol. 51. Springer.

Beibel, M. 1996. A note on ritov's bayes approach to the minimax property of the cusum procedure. *The Annals of Statistics* **24**(4) 1804–1812.

Bertrand, P.R., M. Fhima, A. Guillin. 2011. Off-line detection of multiple change points by the filtered derivative with p-value method. *Sequential Analysis* **30**(2) 172–207.

Brass, W. 1971. On the scale of mortality. *Biological Aspects of Demography* 69–110.

Brémaud, P. 1981. *Point processes and queues, martingale dynamics*. Springer.

Brouhns, N., M. Denuit, J.K. Vermunt. 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* **31**(3) 373–393.

Buhlmann, H. 1970. *Mathematical methods in risk theory*. Springer-Verlag Berlin.

Cairns, A.J.G., D. Blake, K. Dowd, G.D. Coughlan, D. Epstein, A. Ong, I. Balevich. 2009. A quantitative comparison of stochastic mortality models using data from england & wales and the united states. *North American Actuarial Journal* **13** 1–35.

Coughlan, G., M. Khalaf-Allah, Y. Ye, S. Kumar, A.J.G. Cairns, D. Blake, K. Dowd. 2010. Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. *North American Actuarial Journal* .

Cox, D.R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society* 187–220.

Dickson, D.C.M., J.R. Gray. 1984. Exact solutions for ruin probability in the presence of an absorbing upper barrier. *Scandinavian Actuarial Journal* **1984**(3) 174–186.

Fellouris, G., A. Chronopoulou. 2013. Optimal sequential change-detection for fractional diffusion-type equations. *Journal of Applied Probability* **50**(1) 1–20.

Fellouris, G., G.V. Moustakides. 2011. Decentralized sequential hypothesis testing using asynchronous communication. *Information Theory, IEEE Transactions on* **57**(1) 534–548.

Fhima, M., A. Guillin, P.R. Bertrand. 2011. Fast change point analysis on the Hurst index of piecewise fractional Brownian motion. *Preprint* .

Gandy, A., JT Kvaløy, A. Bottle, F. Zhou. 2010. Risk-adjusted monitoring of time to event. *Biometrika* **97**(2) 375–388.

Graversen, S.E., G. Peškir. 1998. Optimal stopping and maximal inequalities for linear diffusions. *Journal of Theoretical Probability* **11**(1) 259–277.

Irbäck, J. 2003. Asymptotic theory for a risk process with a high dividend barrier. *Scandinavian actuarial journal* **2003**(2) 97–118.

Kailath, T., H.V. Poor. 1998. Detection of stochastic processes. *Information Theory, IEEE Transactions on* **44**(6) 2230–2231.

Kella, O., W. Whitt. 1992. Useful martingales for stochastic storage processes with lévy input. *Journal of Applied Probability* 396–403.

Lorden, G. 1971. Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics* **42**(6) 1897–1908.

Lucas, J.M., R.B. Crosier. 1982. Fast initial response for cusum quality-control schemes: give your cusum a head start. *Technometrics* **24**(3) 199–205.

Moustakides, G.V. 1998. Quickest detection of abrupt changes for a class of random processes. *Information Theory, IEEE Transactions on* **44**(5) 1965–1968.

Moustakides, G.V. 1999. Extension of wald's first lemma to markov processes. *Journal of applied probability* **36**(1) 48–59.

Moustakides, G.V. 2004. Optimality of the cusum procedure in continuous time. *The Annals of Statistics* **32**(1) 302–315.

Page, E.S. 1954. Continuous inspection schemes. *Biometrika* **41**(1/2) 100–115.

Poor, H.V. 1998. Quickest detection with exponential penalty for delay. *The Annals of Statistics* **26**(6) 2179–2205.

Ritov, Y. 1990. Decision theoretic optimality of the cusum procedure. *The Annals of Statistics* 1464–1469.

Roberts, S.W. 1966. A comparison of some control chart procedures. *Technometrics* 411–430.

Segerdahl, C.O. 1970. On some distributions in time connected with the collective theory of risk. *Scandinavian Actuarial Journal* **1970**(3-4) 167–192.

Shiryaev, A.N. 1963. On optimum methods in quickest detection problems. *Theory of Probability and its Applications* **8** 22.

Shiryaev, A.N. 1996. Minimax optimality of the method of cumulative sums (cusum) in the case of continuous time. *Russian Mathematical Surveys* **51**(4) 750–751.

Siegmund, D. 1985. *Sequential analysis: tests and confidence intervals*. Springer.

Siegmund, D.O., K.J. Worsley. 1995. Testing for a signal with unknown location and scale in a stationary Gaussian random field. *The Annals of Statistics* **23**(2) 608–639.

Worsley, K.J. 2001. Testing for signals with unknown location and scale in a χ^2 random field, with an application to fMRI. *Advances in Applied Probability* **33**(4) 773–793.