Bulletin Français d'Actuariat manuscript No.
(will be inserted by the editor)

Mortality: a statistical approach to detect model misspecification

Jean-Charles Croix · Frédéric Planchet · Pierre-E. Thérond

Received: date / Accepted: date

Abstract The Solvency 2 advent and the best-estimate methodology in future cash-flows valuation lead insurers to focus particularly on their assumptions. In mortality, hypothesis are critical as insurers use best-estimate laws instead of standard mortality tables. Backtesting methods, i.e. ex-post modeling validation processes, are encouraged by regulators and rise an increasing interest among practitioners and academics. In this paper, we propose a statistical approach (both parametric and non-parametric models compliant) for mortality laws backtesting under model risk. Afterwards, a specification risk is introduced assuming that the mortality law is subject to random variations. Finally, the suitability of the proposed method will be assessed within this framework.

Keywords Solvency 2 · mortality · cusum · detection ·

1 Introduction

The Solvency 2 directive (art. 83, Comparison against experience) imposes that undertakings develop processes to ensure that Best-Estimate calculations and underlying hypotheses are regularly compared against experience. In Life insurance and particularly in annuity computations, mortality models validation and backtesting is of key importance.

In this context, we consider the following simple question: How does an insurer verify that his mortality hypotheses are Best-Estimate ? More precisely, we

Jean-Charles Croix · Frédéric Planchet · Pierre-E. Thérond
ISFA - Université Lyon 1, Laboratoire SAF, 50 avenue Tony Garnier, 69007 Lyon, France

Jean-Charles Croix
E-mail: jean.charles.croix@gmail.com

Frédéric Planchet
Prim'Act - 42 avenue de la Grande Armée - 75017 Paris
E-mail: frederic@planchet.net

Pierre-E. Thérond
Galea & Associés - 12 avenue du Maine - 75015 Paris - France
E-mail: ptherond@galea-associes.eu & pierre@therond.fr

derive testing methodologies to decide whether a given table is likely, according to observations. Indeed, the insurer wants to distinguish sampling variations from misspecification at any age. To do so, a reminder of mortality analysis and models is provided in a first part. The derived statistical models are adequate foundations to develop and support testing processes that detects if prediction errors are the result of sampling variations or an unknown trend. According to these models, a first set of tests with fixed sample sizes are reviewed.

In a second part, the review will be extended to on-line backtesting, which relies on tests with random sample sizes. Indeed, if an insurer repeats fixed tests on a growing set of data (every month for example), the first type error probability converges to one if no corrections are taken on the significance level. This problem is solved using sequential analysis and change-point detection algorithms. Finally, a numerical application is proposed to compare different approaches faced to a simulated misspecification.

2 Mortality models & assumptions

In mortality analysis, life time is considered as a positive random variable T. Considering sufficiently large groups of individuals, mortality risk is assumed mutualized and mathematical models are employed to describe the average behavior of a specific population. Writing S and h the survival and hazard functions respectively, the probability of death between age x and $x+1$ (i.e. at age x) can be expressed as in equation 1 (see Planchet and Thérond (2011)):

$$q_x = P(T \leq x+1 | T > x) = 1 - \frac{S(x+1)}{S(x)} = 1 - \exp\left(-\int_x^{x+1} h(u)du\right). \quad (1)$$

If one wants to predict the number of deaths in a population for a fixed period (without any other causes of population reduction), a minimal segmentation is needed to obtain homogeneity: a simple classifier is age. Under these assumptions, the number of deaths D_x at age x among a population of n_x individuals is a binomial random variable. In a portfolio with p different ages $x \in [x_1, x_p]$, it comes:

$$\forall x \in [x_1, x_p],\ D_x \sim \mathcal{B}(n_x, q_x), \quad (2)$$

in case of annual projections. In the latter, mortality modeling will be summarized in an annual mortality table $q = (q_{x_1}, ..., q_{x_p})$. Furthermore, we will consider observations in monthly requiring a mortality table transformation. If death rates are supposed constant during one year, monthly mortality rates can be derived as follows:

$${}_mq_x = 1 - {}_mp_x = 1 - (1 - q_x)^{\frac{1}{12}}, \quad (3)$$

where ${}_mq_x$ being the desired rate. In the following, all mortality rates are monthly, and the subscript m is omitted. This simple assumption can be refined according to the mortality model implied in table generation. A second assumption in this work is that population renew identically every time-step during analysis.

As a convention in this document, single letters designate vectors over ages (for example, the previously defined q represent a set of p death probabilities), the subscript x is age-specific and the upper-script represents the date of observation

$q^i = (q^i_{x_1}, ..., q^i_{x_p})$. From a statistical view, and whichever the method used to produce the table, it can be considered as a parameter in a parametric model $(\mathcal{Y},\mathcal{P}_Q)$ with $\mathcal{Y}$ the set of all possible observations and $\mathcal{P}_Q$ a family of probability distribution on $\mathcal{Y}$ (see Gourieroux and Monfort (1996) for detailed developments and notations). All previous assumptions can be summarized in the following model:

$$\mathcal{M}_B = \left(\mathcal{Y} = \mathbb{N}^p, \mathcal{P}_Q = \otimes_{x=x_1}^{x_p} \mathcal{B}\left(n_x, q_x\right) | q \in Q\right), \tag{4}$$

with $Q = [0,1]^p$. If this model is well defined, and portfolio sizes are usually large, a Gaussian approximation is often used to simplify computations based on the central limit theorem. Even though this result is asymptotic (i.e. for large n), it's commonly used as the Gaussian law provides ease at use. Furthermore, we'll consider a fixed and known variance-covariance matrix, essentially for simplicity. Finally, we consider the following statistical model:

$$\mathcal{M}_G = \left(\mathcal{Y} = \mathbb{R}^p, \mathcal{P}_Q = \mathcal{N}_p\left(\mu, \Sigma\right) | q \in Q\right), \tag{5}$$

with $\forall x \in [x_1, x_p]$, $\mu_x = n_{x\,m} q_x$ and $(\Sigma)_x = n_{x\,m} q_x^\gamma (1 - {}_m q_x^\gamma)$ a diagonal matrix. In the following, we will consider equivalently deaths and gross mortality rates $\hat{q} = \frac{d}{n}$. Now that our framework for mortality models is defined, we shall present what our backtesting procedure is.

3 Mortality backtesting

Backtesting can be defined as an ex-post model validation method, including two different practices: validation and monitoring. The first aims to validate a mortality table according to a fixed amount of data, while the second allows for continuous treatment. This last aspect can be used to increase power in validation or detect shifts later on.

These problems are usually addressed through decision theory (see Gourieroux and Monfort (1996) or Saporta (2006) for detailed introductions). In our framework, it consists in testing the mean of a Gaussian vector with known variance and detecting any change-point or misspecification. One can find alternative approaches based on different setups (see El Karoui et al (2013) for cox-like models and homogeneous Poisson processes).

Writing q^γ the supposed mortality table and q^0 the real one, the null hypothesis is $H_0 = \{q^\gamma = q^0\}$ against a composite alternative $H_1 = \{q^\gamma \neq q^0\}$. Then tests are defined as couples (ξ_N, N) with N the sample size (possibly random) and ξ_N the associated decision function. All presented procedures are based on likelihood functions, derived from model $\mathcal{M}_G$ but all classical significance tests are applicable. Numerous other tests and change-point procedures can be found elsewhere, especially in change-point detection where the research is still very active in both Frequentist and Bayesian paradigms (see Lai (2001), Tartakovsky and Moustakides (2010b) and Tartakovsky and Moustakides (2010a)). On the other side, sequential alternatives are described in Wald (1947), Ghosh and Sen (1991), Siegmund (1985) and Basseville and Nikiforov (1993).

3.1 Fixed sample tests

Based on the above discussion, we consider fixed sample size tests in this section. In particular, Wald, Score and Likelihood ratio are easily applicable to the previous model and their asymptotic properties (convergence and coverage) are of importance as undertakings usually possess large portfolios. Following Gourieroux and Monfort (1996), we consider the multidimensional constraint $g(q) = q - q^{\gamma}$ which resume the simple hypothesis H_0. In the case of testing the mean of a Gaussian vector, these three tests correspond to the following statistic:

$$\xi = N(\bar{q} - q^{\gamma})^T \Sigma^{-1} (\bar{q} - q^{\gamma}), \tag{6}$$

which is χ^2-distributed under H_0. The associated rejection region W is:

$$W = \{\xi > \chi^2_{1-\alpha}(p)\}, \tag{7}$$

p being the number of ages considered in the portfolio and $\chi^2_{1-\alpha}(p)$ the chi-square quantile with p-degrees of freedom and $1-\alpha$ level. By construction, fixed sample size tests require predefined parameters: a significance level α (or first term error probability) and a predefined sample size N (equivalent to time for periodic observations). In practice, insurers have to define when the test will be conducted: immediately or later with more information ? This decision implies a trade off between fast reaction and power: statistical significance increases with observation as mortality risk. Alternative tests can be found, based on the Standardized Mortality Ratio for example, see Liddell (1984) for example.

3.2 On-line backtesting

In this part, dynamic methods are investigated. The two main related theories are sequential analysis (see Wald (1947)) and change-point detection (Lai (2001) and Basseville and Nikiforov (1993) for detailed presentations and Tartakovsky and Moustakides (2010b) for a more recent review on bayesian technics). Indeed, a simple repetition of previous fixed sample size tests leads to important first type error probability increases.

3.2.1 Sequential Probability Ratio Test (SPRT)

The sequential probability ratio test (SPRT) was first introduced as a test between two simple hypotheses. Constructed on the likelihood ratio Λ_n with n-observations and two thresholds A and B, it's defined as follows:

$$\begin{cases} \textit{reject } H_0, \ if \Lambda_n \geq A \\ \textit{accept } H_0, \ if \Lambda_n \leq B \\ \textit{continue, otherwise} \end{cases} \tag{8}$$

In other terms, the test stops the first time the likelihood ratio leaves the interval $[A, B]$. The corresponding number of observations is called the sample size N and is thus a random variable. Optimality and closure properties are discussed in Wald (1947). Furthermore, the following approximations for α and β holds:

$$\begin{aligned} \alpha &\simeq \frac{1-B}{A-B}, \\ \beta &\simeq \frac{B(A-1)}{A-B}. \end{aligned} \tag{9}$$

These expressions are only approximate due to possible overshoot over boundaries. In case of composite hypotheses, the situation is much more complex and the initial Likelihood Ratio Λ_n must be adapted. Wald (1947) proposed two different solutions. The first is a weighted sequential probability ratio test (WSPRT), obtained specifying prior distribution functions under H_0 and H_1. The second is based on the generalized sequential probability ratio test (GSPRT), using estimators (usually Maximum Likelihood estimators) in place of mixtures. According to Wald (1947), this last version is more difficult to study as the likelihood ratio is no longer a probability distribution (in particular, approximations on error probabilities are not applicable). More recently, Lai (1998) proposed a dynamic boundary for the GSPRT, considering estimators variability.

The main difficulty in the WSPRT design is the choice of an appropriate prior for the parameter q on Q_1 as Q_0 is a singleton. An existing solution is the frequency functions method, based on the likelihood ratio of a sequence of statistics. Using Cox's factorization theorem (in annex), one can reduce composite hypotheses to simple ones using an invariance reduction principle (see Hall et al (1965) for further developments). Applying this method to the gaussian case, Jackson and Bradley (1961) derived χ^2 (and T^2, in case of unknown variance-covariance matrix) sequential probability ratio test, based on homonym statistics. From now on, we apply their result to the previous backtesting problem even though they considered alternative hypotheses of the form: $H_0 = \{\|q^\gamma - q^0\| \leq \epsilon_0\}$ against $H_1 = \{\|q^\gamma - q^0\| \geq \epsilon_1\}$ with $0 \leq \epsilon_0 < \epsilon_1$ implying an indifference region. Depending whether acceptance is needed, one can set $\epsilon_0 = 0$ in the following formula:

$$\ln \Lambda_n = -n\frac{\epsilon_1^2 - \epsilon_0^2}{2} + {}_0F_1\left(\frac{p}{2}, \frac{\epsilon_1^2 n^2 \chi_n^2}{4}\right) - {}_0F_1\left(\frac{p}{2}, \frac{\epsilon_0^2 n^2 \chi_n^2}{4}\right). \tag{10}$$

where ${}_0F_1$ is the generalized hyper-geometric function and $\chi_n^2 = (\bar{q} - q^\gamma)\Sigma^{-1}(\bar{q} - q^\gamma)'$. This result is the ratio between two non-central χ^2 distributions with p-degrees of freedom and respective non-centrality parameter ϵ_0^2 and ϵ_1^2. The choice of A and B is based on Wald's previous approximation which still holds in this case. Unfortunately, there are no practical results to compute the expected sample sizes in this case (but they're are available under i.i.d assumption). One simplification suggested in Jackson and Bradley (1961) is to compute every time step independent statistics using only innovations: the Wald's approximation will hold in despite of a potentially substantial loss of power.

Unlike to the WSPRT, the GSPRT allows a simple test statistic under the initial problem. Here, the difficulty is that Wald's approximation 9 for first and second type errors do not hold anymore. Following developments in Basseville and Nikiforov (1993), it comes:

$$\ln \Lambda_n = \frac{n}{2}\chi_n^2, \tag{11}$$

with $\chi_n^2 = (\bar{q} - q^\gamma)'\Sigma^{-1}(\bar{q} - q^\gamma)$ but we do not have simple expression to derive α and β in order to build a simple procedure to consider whether we have to reject the best estimate assumption.

3.2.2 Quickest detection algorithms

Backtesting can also be interpreted as a change-point detection problem. In this theory, the classical setup is a sequence of random variables distributed under a known distribution f_0, that possibly switches to an alternative and unknown distribution f_1 at an unknown time $\nu \in \mathbb{N}$ (random in Bayesian frameworks and considered equal to ∞ when no changes occur). The objective for change-point detection algorithm τ (defined as a stopping time) is to raise an alarm as quickly as possible when the change occurs, without raising too frequent false alarms. According to Tartakovsky, 4 approaches can be found in the literature: Bayesian (the time of change is random with a specific prior), Generalized Bayesian (improper priors), Multi-cycle procedures and Minimax. Change-point detection is a vast domain and we will focus on frequentist algorithms. Lorden (1971) gave a minimax criterion to compare algorithms in this setup, the essential supremum average detection delay (Notations from Tartakowski):

$$ESADD = \sup_{0\leq\nu<\infty} \operatorname{ess\,sup} E_\nu \left[(\tau-\nu+1)^+|\mathcal{F}_\nu\right]. \tag{12}$$

subject to a constraint of maximal false alarm frequency $E_0(\tau) \geq \lambda$. As a solution to this problem, Page (1954) introduced the Cusum algorithm:

$$\tau = \inf\{n, \Lambda_n - \min_{1\leq j\leq n} \Lambda_j \geq A\} = \inf\{n, \max_{1\leq j\leq n} \Lambda_j^n \geq A\}, \tag{13}$$

where Λ_j^k being the likelihood ratio based on observations j up to k. A recursive version of this algorithm can be found in Lorden (1971) and Basseville and Nikiforov (1993). In his work, Page also pointed out the connection between the Cusum algorithm and Wald's SPRT: the Cusum test can be seen as a set of parallel open-ended SPRTs, a new one starting every period. Writing N_k the sample size of a one-sided open-ended SPRT applied to $\hat{q}_k, \hat{q}_{k+1}, ...$, the Cusum stopping time is $N^\star = \min_{1\leq k\leq n} N_k$.

Equivalently to the SPRT, two solutions are presented to deal with composite hypotheses: the Weighted Cusum $\tilde{\Lambda}$ and the Generalized Likelihood Ratio (GLR) $\hat{\Lambda}$:

$$\begin{aligned}\tilde{\Lambda}_j^k &= \int_{q^1\in\Theta} \frac{\mathcal{L}(\hat{q}^j, ..., \hat{q}^k|q^1)}{\mathcal{L}(\hat{q}^j, ..., \hat{q}^k|q^0)} dF(q^1) \\ &= \frac{\sup_{q^1\in\Theta_1} \mathcal{L}(\hat{q}^j, ..., \hat{q}^k|q^1)}{\mathcal{L}(\hat{q}^j, ..., \hat{q}^k|q^0)}.\end{aligned} \tag{14}$$

Considering previous alternative hypotheses and following Basseville and Nikiforov (1993), two χ^2-Cusum algorithms are available. The first is a direct application (case 3 p.218) of least favorable priors in case of invariant distributions:

$$\ln \tilde{\Lambda}_j^k = -(k-j+1)\frac{\epsilon_1^2}{2} + \ln {}_0F_1\left[\frac{p}{2}, \frac{\epsilon_1^2(k-j+1)^2(\chi_j^k)^2}{4}\right], \tag{15}$$

with $(\chi_j^k)^2 = \left(\bar{q}_j^k - q^\gamma\right)' \Sigma^{-1} \left(\bar{q}_j^k - q^\gamma\right)$. Asymptotic first-order optimality has been proven for the χ^2-Cusum algorithm in multidimensional case (see p.268 in Basseville and Nikiforov (1993)). Introducing $\Delta_0 = E_{\theta_0}(N^\star)$ the mean-time between false alarms and $\Delta_1 = E_{\theta_1}(N^\star)$ the average delay for detection, it comes:

$$\Delta_0 \geq A, \tag{16}$$

from Lorden's theorem (in annex). Furthermore, the χ^2-Cusum algorithm is first-order optimal in that case. The second solution is the Generalized Likelihood Ratio test. According to calculations in Basseville and Nikiforov (1993), the ratio is:

$$\ln \hat{\Lambda}_j^k = \frac{k-j+1}{2}(\chi_j^k)^2. \tag{17}$$

Nevertheless no practical results are available for the GLR.

4 Numerical applications

In this section, we propose a simple numerical illustration to ensure tests efficiency. Tests will first be tested under null hypothesis and then in case of mortality table misspecified. This case will be simulated in a practical method, using a white noise on mortality rates logits. After unbiasing, the table we consider as the real mortality tables q^0 is randomly distributed around the given mortality table q^γ, but equal in mean.

4.1 Misspecification on mortality tables

Specification risk occurs when the given mortality table doesn't fit the real mortality distribution. In this case, if q^0 is the real mortality law, q^γ the model and ϵ the error term it comes:

$$q^0 = f(q^\gamma, \epsilon), \tag{18}$$

where f is an unknown and unobservable function and ϵ a random variable. In this application, our methodology consists in choosing a specific function f and a probability distribution for the error term to produce specification risk. The error term is a controlled Gaussian white noise applied to the pre-defined mortality law logits:

$$\forall x \in [x_1, x_p],\ logit(q_x^0) = logit(q_x^\gamma) + \epsilon_x, \tag{19}$$

with $\epsilon \sim \mathcal{N}_p(0, \sigma Id)$. In other words, the real mortality law is randomly distributed around the pre-defined law q^γ but equal in average ($E(q^0) = q^\gamma$). Thus, the function f is the following:

$$\forall x \in [x_1, x_p],\ q_x^0 = \frac{e^{\epsilon_x} q_x^\gamma}{1 + q_x^\gamma (e^{\epsilon_x} - 1)} - E\left(\frac{e^{\epsilon_x} q_x^\gamma}{1 + q_x^\gamma (e^{\epsilon_x} - 1)} - q^\gamma\right). \tag{20}$$

Finally, an illustration is given of multiple q^0 randomly distributed around q^γ (see figure 1).

Now that specification risk is simulated, the second objective is to find a business interpretation of σ. Indeed, if it's quantitatively defined in previous equations,

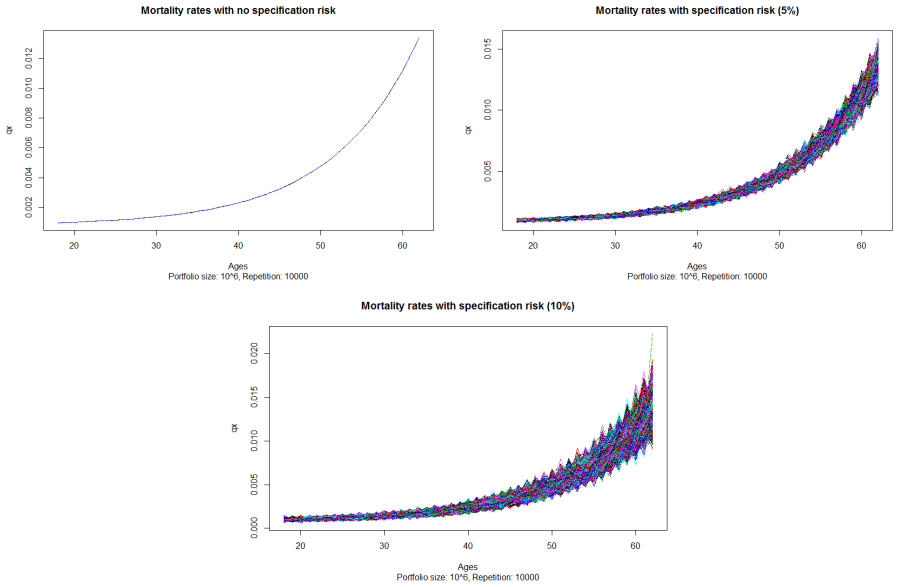

Fig. 1 Example of different levels of specification risk $(0, 5\%, 10\%)$.

Table 1 Correspondence between σ and δ for a 65 years old person and $N = 10^6$.

σ	e	δ
0%	16.21	0.00000
5%	16.34	0.00708
10%	16.48	0.01556
20%	16.75	0.03051
30%	17.00	0.04770
40%	17.23	0.06508

what impact does-it have on real indicators ? For instance, the volatility implied on the remaining life expectancy of a 65-years old male e_{65} is measured as follows:

$$e_{65} = \frac{1}{S(65)} \sum_{j=66}^{120} S(j), \tag{21}$$

with $S(x) = \prod_{i=1}^{x-1} (1 - q_i)$ the discrete survival function. Considering e_{65} as a function of ϵ, here is a measure of the deviation of e_{65}:

$$\delta = \frac{q_{95\%}(e_{65}) - E(e_{65})}{E(e_{65})}. \tag{22}$$

The following table 1 shows correspondence between remaining life expectancy volatility δ and σ (see annex for detailed computation of δ).

4.2 Data simulation and portfolio structure

The test methodology consists in setting first q^{γ} (in our example, it has been adjusted on the French regulatory mortality table TH00-02). Then, for each simula-

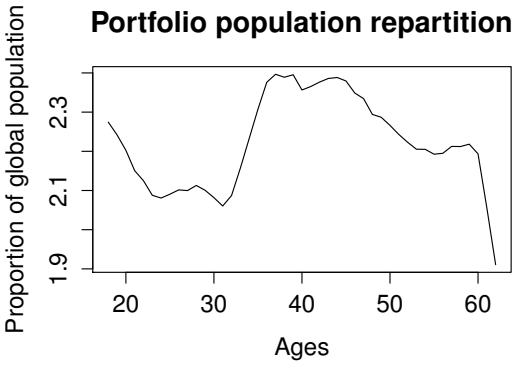

Fig. 2 Population repartition over ages in proportions.

Table 2 Tests results: $\alpha = 5\%$, $\sigma = 0\%$, 1000 simulations, 60 months

	R	$E(N)$
χ^2-SPRT	0.038	11.89
GLR-SPRT	0.022	19.90
χ^2-CUSUM	0.381	29.52
GLR	0.380	28.79
χ^2	0.056	12.00

tion, a noise is simulated and applied to obtain q^0. From that, deaths are generated every month and tests conducted. The portfolio population is based on the French Insee demographic structure (table RP2009) and includes people between 18 and 62 years-old (see figure 2) for a total of 10^6 individuals.

Numerical results are available in Table 2 to Table 7, and should be read as follows: R is the rejection rate and $E(N)$ is the observed expected sample size (ignoring non rejections). In particular, the χ^2 test is conducted only once on the 12-th month, thus $E(N) = 12$ for this test. Secondly, the GLR and GLR-SPRT tests are set-up with a boundary of $2A$ instead of A to take in account estimators deviation.

In table 2, 3 and 7, tests are driven under H_0 for two different levels of α. We can observe that χ^2, χ^2-SPRT and GLR-SPRT are controlled in terms of first-type error probabilities. Furthermore, the $\chi^2 - SPRT$ has a lower rejection rate in comparison with fixed test and a close expected sample size, even if the test continuous beyond 12 months (in particular, if the test was running only on 12 months, the gap would be even higher). In comparison, the rejection rates of change-point detection algorithms are far higher, even if the rejection appear to be later on (higher expected sample sizes). Finally, in case of H_0, we prefer the χ^2-SPRT which achieve better results.

Tables 4, 5 and 6 show tests behavior for non-null noise. As expected, rejection rates increase with noise and the overall expected sample size reduces. In particular, GLR solutions achieve the lowest $E(N)$ while they remain close to χ^2 CUSUM and SPRT performances.

Table 3 Tests results: $\alpha = 1\%$, $\sigma = 0\%$, 1000 simulations, 60 months

	R	$E(N)$
χ^2-SPRT	0.009	12.56
GLR-SPRT	0	/
χ^2-CUSUM	0.105	33.26
GLR	0	/
χ^2	0.009	12.00

Table 4 Tests results: $\alpha = 5\%$, $\sigma = 10\%$, 1000 simulations, 60 months

	R	$E(N)$
χ^2-SPRT	1	10.01
GLR-SPRT	1	11.50
χ^2-CUSUM	1	9.49
GLR	1	10.60
χ^2	0.91	12.00

Table 5 Tests results: $\alpha = 5\%$, $\sigma = 20\%$, 1000 simulations, 60 months

	R	$E(N)$
χ^2-SPRT	1	3.90
GLR-SPRT	1	3.31
χ^2-CUSUM	1	3.90
GLR	1	3.27
χ^2	1	12.00

Table 6 Tests results: $\alpha = 5\%$, $\sigma = 10\%$, 1000 simulations, 12 months

	R	$E(N)$
χ^2-SPRT	0.82	8.11
GLR-SPRT	0.66	8.46
χ^2-CUSUM	0.85	8.02
GLR	0.72	8.15
χ^2	0.93	12.00

5 Conclusion

In conclusion of this work, we have presented how statistical modelling, through fixed sample size tests, sequential analysis and change-point detection algorithms can insure an effective mortality backtesting. Far from being exhaustive, our approach provides fast and simple methods to follow continuously, with controlled first-type error probability and with an acceptable power mortality risk. Indeed, empirical results shows a superior power for fixed sample size tests but they don't provide a suitable practical framework. Furthermore, change-point detection algo-

Table 7 Tests results: $\alpha = 5\%$, $\sigma = 0\%$, 1000 simulations, 12 months

	R	$E(N)$
χ^2-SPRT	0.03	8.32
GLR-SPRT	0.01	4.00
χ^2-CUSUM	0.05	9.01
GLR	0.09	6.66
χ^2	0.06	12.00

rithms can also be applied to detect shifts in mortality trends. Finally, we believe that sequential analysis and change-point detection processes can be applied to more complex situations, including disability and multiple other causes. We have to insist that the presented procedures lead to a symmetric appreciation of the tested mortality assumptions. This could and should lead to different consequences depending on whether this lead to an overestimation or underestimation of the predicted risks. Using such techniques enable to get a quantitative appreciation in order to accompany an expert's judgement on the reliability of the mortality assumption.

A Cox's theorem

Let $x = [x_1, ..., x_n]$ be random variables whose probability density function (p.d.f.) depends on unknown parameters $\theta_1, ..., \theta_p$. The x_i themselves may be vectors. Suppose that:

- (i) $t_1, ..., t_n$ are a functionally independent jointly sufficient set of estimators for $\theta_1, ..., \theta_p$,
- (ii) the distribution of t_1 involves θ_1 but not θ_2,...,θ_p,
- (iii) $u_1, ..., u_m$ are functions of x functionally independent of each other and $t_1, ..., t_p$,
- (iv) there exists a set S of transformations of $x = [x_1, ..., x_n]$ into $x^* = [x_1^*, ..., x_n^*]$ such that
  - (a) $t_1, u_1, ..., u_m$ are unchanged by all transformations in S,
  - (b) the transformation of $t_2, ..., t_p$ into $t_2^*, ..., t_p^*$ is one-to-one,
  - (c) if $T_1, ..., T_p$ and $T_2^*, ..., T_p^*$ are two set of values of $t_2, ..., t_p$ each having non-zero probability density under at least one of the distributions of x, then there exists a transformation in S such that if $t_2 = T_2, ..., t_p = T_p$, then $t_2^* = T_2^*, ..., t_p^* = T_p^*$.

Then the joint p.d.f. of $t_1, u_1, ..., u_m$ factorizes into

$$g(t_1, \theta_1) l(u_1, ..., u_m, t_1), \tag{23}$$

where g is the p.d.f. of t_1 and l doesn't involve θ_1.

B Lorden's theorem

Let N be a stopping time (or equivalently a sample size) with respect to $y_1, y_2, ...$ such that

$$P_0(N < \infty) \leq \alpha. \tag{24}$$

For $k = 1, 2, ...$, let N_k be the stopping time obtained by applying N to $y_k, y_{k+1}, ...$ Define the extended stopping time $\tau = \min(k, N_k)$, then:

$$\begin{aligned} \Delta_0(\tau) &\geq \frac{1}{\alpha}, \\ \Delta_1(\tau) &\leq E_1(N). \end{aligned} \tag{25}$$

Acknowledgements This work has been supported by the BNP Paribas Cardif Chair "Management de la modélisation". The authors would like to thank Yahia Salhi for his very hepful advice about detection problems.

References

Basseville M, Nikiforov IV (1993) Detection of Abrupt Changes: Theory and Applications. Prentice-Hall

El Karoui N, Loisel S, Salhi Y, Mazza C (2013) Fast change detection on proportionnal two-population hazard rates. Working paper
Ghosh BK, Sen PK (1991) Handbook of Sequential Analysis. CRC Press
Gourieroux C, Monfort A (1996) Statistique et Modéles économétriques. Economica
Hall J W, Wijsman A R, Ghosh JK (1965) The relationship between sufficiency and invariance with applications in sequential analysis. The Annals of Mathematical Statistics 36:575–614
Jackson E J, Bradley A R (1961) Sequential chi-2 and t-2 tests. JST pp 1063–1077
Lai L T (1998) Nearly optimal sequential tests of composite hypotheses. The Annals of Statistics 16:856–886
Lai L T (2001) Sequential analysis: Some classical problems and new challenges. Statistica Sinica pp 303–408
Liddell K F D (1984) Simple exact analysis of the standardised mortality ratio. Journal of Epidemiology and Community Health 38:85–88
Lorden G (1971) Procedures for reacting to a change in distribution. The Annals of Mathematical Statistics 42:1897–1908
Page S E (1954) Continuous inspection schemes. Biometrika 1-2:100–115
Planchet F, Thérond P (2011) Modelisation statistique des phénomènes de durée : Applications actuarielles. Economica
Saporta G (2006) Probabilités, analyses de données et statistiques. Technip
Siegmund D (1985) Sequential analysis: Tests and Confidence Intervals. Springer Verlag
Tartakovsky AG, Moustakides GV (2010a) Discussion on "quickest detection problems: Fifty years later" by albert n. shiryaev. Sequential Analysis 29:386–393
Tartakovsky AG, Moustakides GV (2010b) State-of-the-art in bayesian change-point detection. Sequential Analysis 29:125–145
Wald A (1947) Sequential Analysis. Dover Phoenix Editions