Mortality: A statistical approach to detect model misspecification

Life colloquium presentation

Jean-Charles CROIX

ISFA, Université Claude Bernard
Université Lyon 1
BNP Paribas Cardif

June 2013, the 26th

joint work with F. Planchet and P. Thérond

Summary

Motivation

- Regulatory requirement: Solvency II
  - Best-estimate evaluation
  - Art. 83, comparison against experience
- Risk management good practice:
  - Primary risk factor in annuity business
  - Cash-flow projection model input
- Validation of complex data manipulation
- Simplicity for practical applications

Mathematical foundations

The underlying model is classical in mortality analysis, considering life-time as a random variable T. Under usual notations:

Annual death probability at age x

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

S, h being respectively the survival and hazard functions. Thus, considering an annual step and a population of n_x individuals, deaths are independent binomial random variables (no other exit causes are taken in account):

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

Practical approximations

In our practical application, we suppose a given annual mortality table q^{γ}. As we consider monthly observations, we apply the following approximation for analysis (piecewise constant hazard function):

$$ {}_{m}q_{x} = 1 - (1 - q_{x})^{\frac{1}{12}} \tag{3} $$

which leads to the following statistical model, assuming a normal approximation and an exact renewal of population under risk each time steps:

Statistical model

$$ \mathcal{M}_{G} = (\mathcal{Y} = \mathbb{N}^{p}, \mathcal{P}_{Q} = \mathcal{N}_{p}(n_{m}q, \Sigma)|_{m}q \in Q) \tag{4} $$

with $Q = [0, 1]^{p}$ the parameter space, $\mathcal{N}_{p}$ the p-dimensionnal normal law and Σ the variance-covariance matrix.

Mortality Backtesting: Definition

A **backtest** is an **ex-post** model validation method. In mortality analysis, it corresponds to tables **validation** and **monitoring**. Let q^γ be the mortality table to test and q^0 the real set of death probabilities. A natural criterion for comparison is to decide between the two following hypotheses:

$$\begin{aligned} H_0 &= \{(q^\gamma, q^0) \in Q^2,\ \|q^\gamma - q^0\| = 0\} = \{q^\gamma \in Q_0\}, \\ H_1 &= \{(q^\gamma, q^0) \in Q^2,\ \|q^\gamma - q^0\| > 0\} = \{q^\gamma \in Q_1\}, \end{aligned} \tag{5}$$

$\|.\|$ a convenient norm and $Q_0 \cap Q_1 = \emptyset$. Backtesting procedures will be compared according to first and second type errors probabilities, α and β. Multiple purposes can be achieved through backtesting:

- Hypothesis validation: Mortality table validation (Statistical and sequential tests).
- Hypothesis monitoring: Continuous control (Change-point detection algorithms).

Backtesting: Validation

Consider a **fixed** number of vectorial observations N: $d^N = (d_1, ...d_N)$ or equivalently $\hat{q}^N = (\hat{q}_1, ..., \hat{q}_N)$. Based on the statistical model $\mathcal{M}_G$, it's straightforward to derive likelihood based tests:

Likelihood based tests, Gourieroux and Monfort (1996)

Wald, Score and Likelihood Ratio tests are defined respectively as follows:

$$\begin{aligned} \xi^W &= (\hat{q} - q^\gamma)^t \mathcal{I}(\hat{q})(\hat{q} - q^\gamma), \\ \xi^S &= \frac{\partial \ln \mathcal{L}(D, q^\gamma)}{\partial q}^t \mathcal{I}^{-1}(q^\gamma) \frac{\partial \ln \mathcal{L}(D, q^\gamma)}{\partial q}, \\ \xi^R &= 2(\ln \mathcal{L}(D, \hat{q}) - \ln \mathcal{L}(D, q^\gamma)). \end{aligned} \tag{6}$$

In case of one observation, the rejection region being $\{\xi \geq \chi^2_{(1-\alpha)}(p)\}$ in case of a test of level α.

These tests are asymptotically convergent, of coverage 1, equivalent and χ^2 distributed.

Backtesting: Validation

Consider N observations and $\bar{q} = \frac{1}{N}\sum_{i=1}^{N} \hat{q}^i$ its mean vector.

Likelihood based tests, applications in the binomial case

The Wald, Score and Likelihood Ratio tests are defined respectively as follows:

$$\begin{aligned}
\xi^W &= \sum_{x=x_1}^{x_p} \frac{n_x(\bar{q}_x - q_x^\gamma)^2}{\bar{q}_x(1-\bar{q}_x)} \\
\xi^S &= \sum_{x=x_1}^{x_p} \frac{n_x(\bar{q}_x - q_x^\gamma)^2}{q_x^\gamma(1-q_x^\gamma)} \\
\xi^R &= \sum_{x=x_1}^{x_p} D_x \ln\left(\frac{\bar{q}_x}{q_x^\gamma}\right) + (n_x - D_x)\ln\left(\frac{1-\bar{q}_x}{1-q_x^\gamma}\right)
\end{aligned} \tag{7}$$

Very good results for large portfolios (asymptotic in n_x) but unknown β.

Backtesting: Validation

The SPRT is designed to accept additional observations as long as statistical significance (in both α and β) can't be assessed. Thus, the sample size N (number of required observations before decision) is a random variable. This test requires an indecision region.

Sequential Probability Ratio Test (SPRT), Wald (1947)

The sequential probability ratio test is defined as follows:

- reject H_0, if $\Lambda_n \geq A$
- accept H_0, if $\Lambda_n \leq B$
- continue otherwise

with Λ_n a likelihood ratio with n observations and $A > B$ two constants.

$$N = \inf\{n, \Lambda_n \notin [B, A]\} \tag{8}$$

Difficulty: Choose likelihood ratio Λ_n, A and B (Possibly 0).

Backtesting: Validation

SPRT Properties, Wald (1947)

- The SPRT test is optimal for simple hypotheses.
- Wald's approximations for closed tests i.e. $P(N < \infty) = 1$:
  - $\alpha \simeq \frac{1-B}{A-B}$
  - $\beta \simeq \frac{B(A-1)}{A-B}$
- In addition, if observations are i.i.d, $E(N)$ is available under both hypotheses.

Wald proposes two methods for composite hypotheses, based on a weighted likelihood ratio $\tilde{\Lambda}_n$ or a maximum likelihood estimator $\hat{\Lambda}_n$:

$$\tilde{\Lambda}_n = \frac{\int_{Q_1} \mathcal{L}(\hat{q}_1, ..., \hat{q}_n, q) dF(q)}{\int_{Q_0} \mathcal{L}(\hat{q}_1, ..., \hat{q}_n, q) dF(q)} \text{ and } \hat{\Lambda}_n = \frac{\sup_{Q_1} \mathcal{L}(\hat{q}_1, ..., \hat{q}_n, q)}{\sup_{Q_0} \mathcal{L}(\hat{q}_1, ..., \hat{q}_n, q)} \tag{9}$$

Difficulty: choice of priors F, unknown error probabilities for $\hat{\Lambda}_n$ (not a distribution).

Backtesting: Validation

Prior choice: Least favorable functions, Basseville and Nikiforov (1993). The method of *frequency functions* considers observations as sequence of test-statistics and is supported by the following theorem (Invariant Sequential tests):

Cox's theorem, related from Jackson and Bradley (1961)

Let $x = [x_1, ..., x_n]$ be random varibables 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 $t_1, ..., t_p$,

Backtesting: Validation

Cox's theorem (suite)

- (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

$$g(t_1, \theta_1) l(u_1, ..., u_m, t_1) \tag{10}$$

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

Backtesting: Validation

Based on slightly different hypotheses (no contact between Q_0 and Q_1), Jackson and Bradley first derived χ^2-SPRT.

Application of Cox's theorem, from Jackson and Bradley (1961)

Considering previous problem of testing:

- $H_0 = \{q^0 \in Q, (q^0 - q^\gamma)\Sigma^{-1}(q^0 - q^\gamma)' = 0\}$
- $H_1 = \{q^0 \in Q, (q^0 - q^\gamma)\Sigma^{-1}(q^0 - q^\gamma)' = \lambda^2\}$

and the χ^2 statistic, it comes:

$$\tilde{\Lambda}_n = \exp\left(-n\frac{\lambda^2}{2}\right) {}_0F_1\left(\frac{p}{2}, \frac{n^2\lambda^2\chi_n^2}{4}\right) \tag{11}$$

with $\chi_n^2 = (\bar{q} - q^\gamma)\Sigma^{-1}(\bar{q} - q^\gamma)'$ and ${}_0F_1$ the generalized hypergeometric function.

Difficulty: Choice of λ in Life insurance context.

Backtesting: Monitoring

From now on, the mortality table q^γ is supposed to be true in the beginning and the objective will be to continue approval or detect shifts.

Change-point detection problem

Consider a sequence of independent random variables $(Y_n)_{n\in\mathbb{N}}$ distributed under $\mathbb{P}_0$ until an unknown observation Y_ν (ν being possibly infinite) and then $\mathbb{P}_1$. Let τ be a stopping time associated to a detection algorithm, the objective is to minimize the following criterion from Lorden (1971):

$$\sup_{\nu\geq 1} E_\nu\left(\tau-\nu+1|\tau\geq\nu\right) \tag{12}$$

subject to $E_\infty(\tau)\geq$ large constant.

Mutliple technics exist: Shewart control charts, Geometric moving averages and other statistical process control (SPC), sequential tests, repeated fixed sample size tests...

Backtesting: Monitoring

Change-point detection algorithms seek to minimize the worst mean delay for detection for a large mean-time between false alarms. Consider a change-point detection algorithm represented by its stopping time τ:

Mean-time between false alarms

$$\bar{\tau} = E_{\infty}(\tau) \tag{13}$$

Worst mean delay for detection

$$\tau^{*} = \sup_{\nu} ess\sup E_{\nu}\left(\tau - \nu + 1 | \tau \geq \nu, Y_1, ..., Y_n\right) \tag{14}$$

Optimality of change-point algorithm

An optimal algorithm minimizes the Worst mean delay for a predefined mean-time between false alarm (Asymptotically optimal when $\bar{\tau} \to \infty$).

Backtesting: Monitoring

Lorden's theorem, Lorden (1971)

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

$$P_0(N < \infty) \leq \alpha \tag{15}$$

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:

$$\begin{aligned} \bar{\tau} &\geq \frac{1}{\alpha} \\ \tau^* &\leq E_1(N) \end{aligned} \tag{16}$$

This result gives a lower boundary for mean-time between false alarm, useful in practice.

Backtesting: Monitoring

Cumulative Sums (Cusum), Page (1954)

The Cusum algorithm is a stopping time defined as follows:

$$\tau = \inf\{n, \Lambda_n - \min_{1 \leq j \leq n} \Lambda_j \geq A\} = \inf\{n, \max_{1 \leq j \leq n} \Lambda_j^n \geq A\} \tag{17}$$

Page showed that this method can be seen as a repetition of one-sided SPRT ($B \to 0$), each starting at a different date.

Relation between Cusum and SPRT

Noting N_k the sample size associated with the SPRT starting from observation k and τ the Cusum stopping time, Page showed the following connection:

$$\tau = \min(k, N_k) \tag{18}$$

Backtesting: Monitoring

As for the SPRT test, two different methods can be applied to composite hypotheses:

Weighted Cusum

$$\tilde{\Lambda}_j^k = \frac{\int_{q \in Q_1} \mathcal{L}(\hat{q}_j, ..., \hat{q}_k, q) dF_1(q)}{\int_{q \in Q_0} \mathcal{L}(\hat{q}_j, ..., \hat{q}_k, q) dF_0(q)} \tag{19}$$

Generalized likelihood ratio

$$\hat{\Lambda}_j^k = \frac{\sup_{q \in Q_1} \mathcal{L}(\hat{q}_j, ..., \hat{q}_k, q)}{\sup_{q \in Q_0} \mathcal{L}(\hat{q}_j, ..., \hat{q}_k, q)} \tag{20}$$

Backtesting: Monitoring

χ^2 Cusum algorithm, Basseville and Nikiforov (1993)

$$\tilde{\Lambda}_j^k = \exp\left(-(k-j+1)\frac{\lambda^2}{2}\right) {}_0F_1\left[\frac{p}{2}, \frac{\lambda^2(k-j+1)^2\chi_j^k}{4}\right] \tag{21}$$

GLR, Basseville and Nikiforov (1993)

$$\hat{\Lambda}_j^k = \exp\left((k-j+1)\left(\lambda\chi_j^k - \frac{\lambda^2}{2}\right)\right) \tag{22}$$

with $\chi_j^k = (\bar{q}_j^k - q^\gamma)\Sigma^{-1}(\bar{q}_j^k - q^\gamma)'$ and $\bar{q}_j^k = \frac{1}{k-j+1}\sum_{i=j}^k \hat{q}^i$

Backtesting: Application

Simulation of misspecification

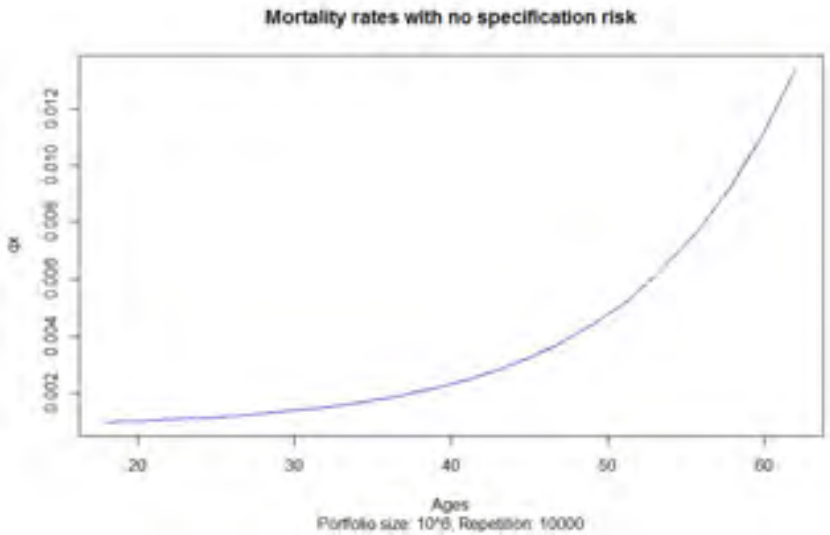
In order to generate model misspecification, we'll consider the following transformation. The first application consist in a gaussian noise applied on mortality logits.

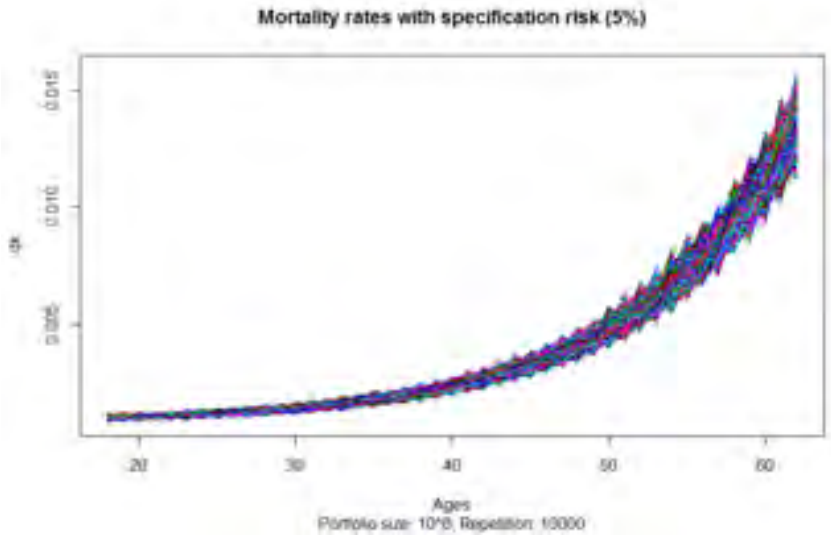
$$\forall x \in [x_1, x_p],\ logit(q_x^0) = logit(q_x^\gamma) + \epsilon_x \tag{23}$$

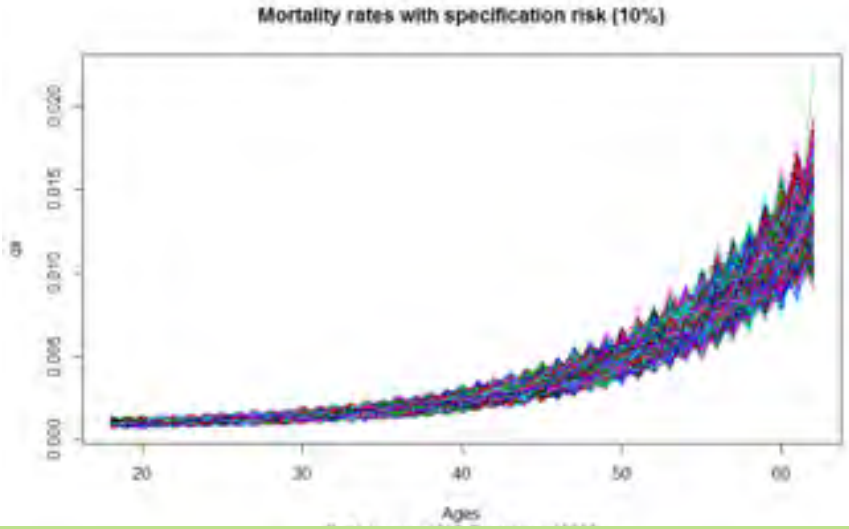
The second step in unbiasing the resulting mortality rates, which finally leads to the following relation:

$$\forall x \in [x_1, x_p],\ q_x^0 = \frac{e^{\epsilon_x} q_x^\gamma}{1 + q_x^\gamma (e^{\epsilon_x} - 1)} - E\left(\frac{e^{\epsilon_x} q_x^\gamma}{1 + q_x^\gamma (e^{\epsilon_x} - 1)} - q^\gamma \right). \tag{24}$$

Backtesting: Application







Backtesting: Application

The following table presents equivalence between noise volatility on logits and incertitude on life remaining expectancy for a 65 years old men.

$$\delta = \frac{q_{95\%}(e_{65}) - E(e_{65})}{E(e_{65})}. \tag{25}$$

Table : 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

Backtesting: Application

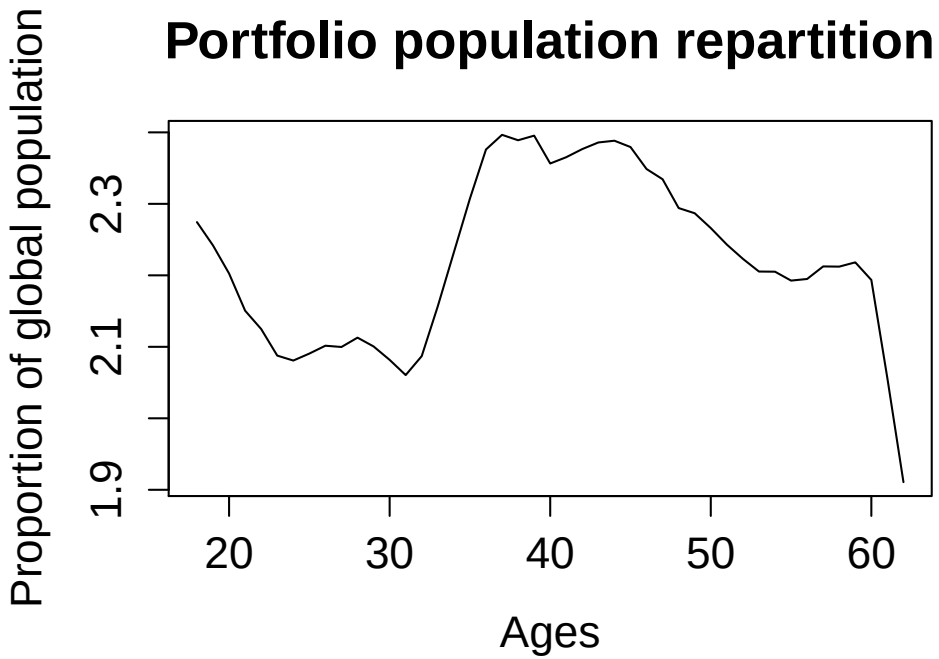


Figure : Population repartition over ages in proportions. source: Insee RP09, french demography.

Backtesting: Application

Table : Test results with $n = 10^6$, $\alpha = 5\%$, $nMonths = 60$, $\lambda = 1.6$ and $\sigma = 0\%$.

σ	$E(N \mid N \leq nMonths)$	$P(reject)$
χ^2-SPRT	12.46	0.032
χ^2-Cusum	31.84	0.385
ξ^S	12 (fixed)	0.057

Table : Test results with $n = 10^6$, $\alpha = 1\%$, $nMonths = 60$, $\lambda = 1.6$ and $\sigma = 0\%$.

σ	$E(N \mid N \leq nMonths)$	$P(reject)$
χ^2-SPRT	16.33	0.003
χ^2-Cusum	37.07	0.082
ξ^S	12 (fixed)	0.007

Backtesting: Application

Table : Test results with $n = 10^6$, $\alpha = 5\%$, $nMonths = 60$, $\lambda = 1.6$ and $\sigma = 10\%$.

σ	$E(N \mid N \leq nMonths)$	$P(reject)$
χ^2-SPRT	9.613	1
χ^2-Cusum	9.114	1
ξ^S	12 (fixed)	0.927

Table : Test results with $n = 10^6$, $\alpha = 5\%$, $nMonths = 60$, $\lambda = 1.6$ and $\sigma = 20\%$.

σ	$E(N \mid N \leq nMonths)$	$P(reject)$
χ^2-SPRT	3.92	1
χ^2-Cusum	3.916	1
ξ^S	12 (fixed)	1

Conclusion & Orientation

- Insurance applications require composite hypotheses tests, complex setups,
  - Asymptotic optimality results
  - Difficult parameters tuning (λ,...), prior on λ
- Practical advantage of sequential tests:
  - Good power in comparison with fixed sample tests,
  - Controlled error probabilities.
- GLR method: Numerical difficulties
- Include alternative outgoes: Censure, Disability
- Derive tests without parameter indifference region

Main References & Bibliography

Croix J-C., Planchet F. and Therond P. (Working paper), Mortality: A statistical approach to detect model misspecification, available on http://hal.archives-ouvertes.fr/hal-00839339
El Karoui N., Loisel S., Sahli Y and Mazza C. (Working paper), Fast change detection on proportional two-population hazard rates
Wald A. (1947), Sequential Analysis, Dover phoenix editions
Lorden G. (1971), Procedures for reacting to a change in distribution, The Annals of Mathematic statistics, vol. 42, N.6, 1897-1908
Basseville M. and Nikiforov I.V. (1993), Detection of abrupt changes: Theory and Applications, Prentice-Hall.