Constructing Entity Specific Projected Mortality Table: Adjustment to a Reference

Julien TOMAS [α] [*]   Frédéric PLANCHET [α] [†]

[α] *ISFA - Laboratoire SAF* [‡]

May 28, 2014 - Version 1.3

Abstract

This article presents an operational framework for constructing and validating projected mortality table specific to an insurer. We describe several methods of increasing complexity and the process of validation allowing an insurer to adjust a reference to get closer to a best estimate assessment of its mortality and longevity risk. They provide to the insurer some latitude of choice while preserving simplicity of implementation for the basic methodology. The methodologies articulate around an iterative procedure allowing to choose parsimoniously the most satisfying one. The process of validation is assessed on three levels. This concerns the proximity between the observations and the model, the regularity of the fit as well as the plausibility and consistency of the mortality trends. Finally, the procedure is illustrated from experience data originating from a French Insurance portfolio.

Keywords. Mortality modelling, Validation, Prospective morality table, Relational models, Local likelihood, Generalized linear models, Life insurance, Graduation, Extrapolation.

Résumé

Cet article fournit un cadre opérationnel concernant la construction et la validation de tables de mortalité prospectives spécifiques à une entité. A partir d'un historique restreint et de tables de mortalité prospectives *best estimate* de place, nous décrivons les méthodologies de positionnement, ainsi que le processus de validation, permettant à un organisme donné d'ajuster la référence pour se rapprocher d'une vision *best estimate* de ses engagements. Plusieurs méthodologies de complexité croissante sont prtésentées. Elles laissent aux organismes une certaine latitude de choix, tout en préservant une grande simplicité de mise en œuvre pour la méthodologie de base. La validation est apprtéhendée sur trois niveaux. Celle-ci concerne la proximité entre les observations et le modèle, la régularité de l'ajustement ainsi que la plausibilité et la cohérence des tendances d'évolution de la mortalité. Enfin, nous illustrons la procédure à partir de données d'expérience originaires d'un portefeuille français d'assurance.

Mots-clés. Modèlisation de la mortalité, Validation, Table de mortalité prospective, Modèles relationnels, Vraisemblance locale, Modèles linéaires généralisés, Assurance vie, Graduation, Extrapolation.

---

[*]Contact: julien.tomas@univ-lyon.fr.

[†]Contact: frederic.planchet@univ-lyon1.fr.

[‡]Institut de Science Financière et d'Assurances - Université Claude Bernard Lyon 1 - 50 Avenue Tony Garnier - 69366 Lyon - France

Contents

1 Introduction

In this article, we present an operational framework for constructing and validating projected mortality tables specific to an insurer. The material is based on studies carried out by Institut des Actuaires, available publicly at `www.ressources-actuarielles.net/gtmortalite`. These researches have been conducted with in the aim of providing to French insurance companies methodologies to take into account their own mortality experience for the computation of their best estimate reserves.

Recently, Hunt and Blake (2013) have proposed a general procedure for constructing mortality models using a combination of a toolkit of functions and expert judgement. They implement an iterative process to identifies every significant demographic feature in the data and apply expert judgement to choose a particular parametric form to represent it.

Here, we propose an operational procedure in a more restrictive framework. We describe several methodologies of increasing complexity and the process of validation allowing an insurer to adjust a baseline mortality to get closer to a best estimate assessment of its mortality and longevity risks.

The techniques proposed are based only on the two following elements:

i. A baseline mortality. In this article, we use the mortality tables presented in the technical report Tomas and Planchet (2013b), constructed for the French insurance market,

ii. Data line by line originating from a portfolio provided by an insurer.

Various methods of increasing complexity are presented. They allow the insurer some latitude of choice while preserving simplicity of implementation for the basic methodology. In addition they provide a simple adjustment without any intervention of an expert.

The simplest approach is the application of a single factor of reduction / increase to the probabilities of death of the baseline mortality. In practice, this coefficient is the Standardized Mortality Ratio (SMR) of the population considered computed on a global level.

The second method is a semi-parametric Brass-type relational model. This model implies that the differences between the observed and baseline mortality can be represented linearly with two parameters.

For the third method, we consider a Poisson generalized linear model including the baseline mortality as a covariate and allowing interactions with age and calendar year.

The fourth method includes, in a first step, a non-parametric smoothing of the periodic table and, in a second step, the application of the rates of mortality improvement derived from the baseline mortality.

The validation is assessed on three levels. The first level concerns the proximity between the observations and the model. It is assessed by the likelihood ratio test, the SMR test and the Wilcoxon Matched-Pairs Signed-Ranks test. In complement, the validation of the fit involves graphical diagnostics such as the analysis of the response, Pearson and deviance residuals as well as the comparison between the predicted and observed mortality by attained age and calendar year.

The second level involves the regularity of the fit. It is assessed by the runs test and signs test.

Finally, the third level covers the plausibility and consistency of the mortality trends. It is evaluated by singles indices summarizing the lifetime probability distribution for different cohorts at several ages. Moreover, it involves graphical diagnostics assessing the consistency of the observed and projected life expectancy. Additionally, if we have at our disposal the male and female mortality, we can compare the improvement and judge the plausibility of the common evolution of the mortality of the two genders.

We ask the question where the data are originating from and based on this knowledge, what mixture of biological factors, medical advances and environmental changes would have to happen to cause this particular set of forecasts.

The methods articulate around an iterative procedure allowing to compare them and to choose parsimoniously the most satisfying one. The procedure is as follow. We start with the first method, if the criteria corresponding to the first level of validation are not satisfied, we switch to the second method and it is useless to continue the validation with the first method. If the criteria corresponding to the first level are satisfied, we continue the validation with the second level criteria. We can also turn to the next method to improve the fit at a cost of a somewhat greater complexity without degrading the results of the first level criteria. The process of validation is summarize in Figure 1.
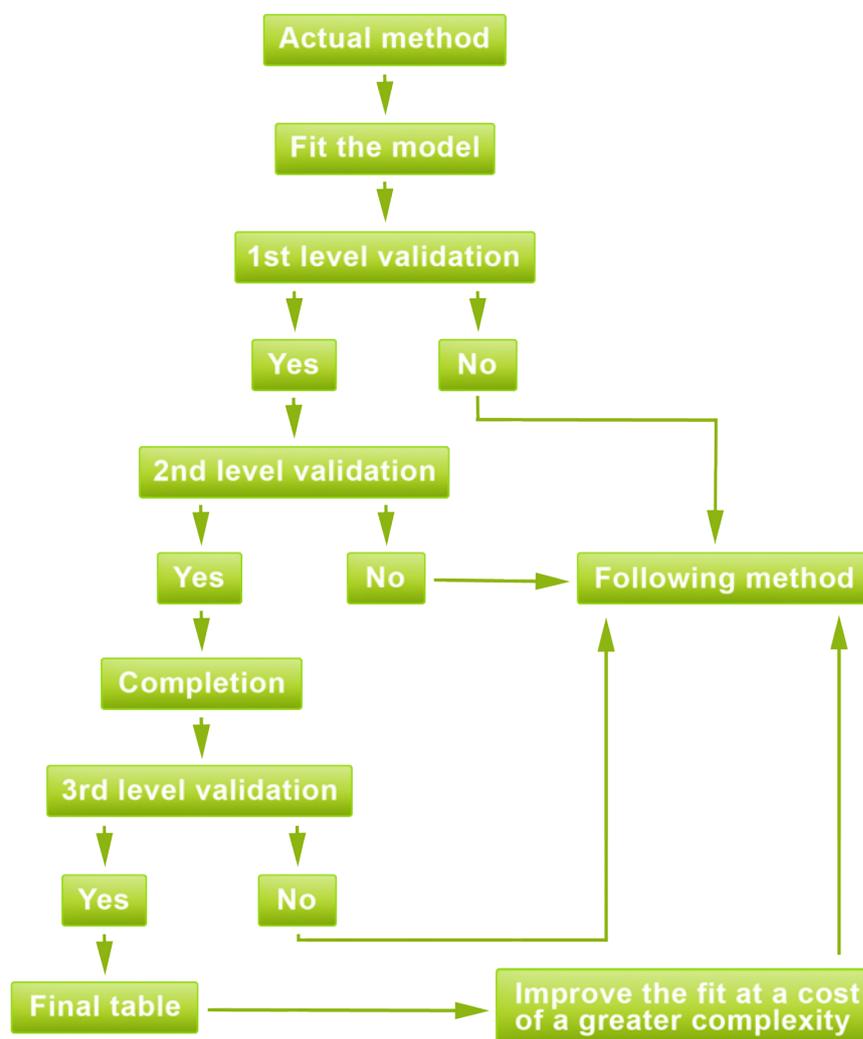
Figure 1: The process of validation.

In addition, a practical guide implementing the methodologies, based on the software R, R Development Core Team (2014) and on the package ELT, see Tomas and Planchet (2014), is available at www.ressources-actuarielles.net/gtmortalite.

This article is organized as follows. Section 2 has still an introductory purpose. It specifies the notation and assumptions used in the following. Section 3 assesses the methodological aspects of the four methods considered. The three levels of validation are covered in Section 4. Section 5 presents an application of ours methodologies with data originating from a French insurance portfolio. Finally, some remarks in Section 6 conclude the paper.

2   Notation and assumption

We analyze the mortality as a function of both the attained age x and the calendar year t. The number of individuals at attained age x during calendar year t is denoted by $L_{x,t}$ and $D_{x,t}$ represents the number of deaths recorded from an exposure-to-risk $E_{x,t}$ that measures the time during individuals are exposed to the risk of dying. We suppose that we have data line by line originating from a portfolio. To each of the observations i, we associate the dummy variable δ_i indicating if the individual i dies or not over one year, i.e. in the age-band $[x, x+1]$,

$$\delta_i = \begin{cases} 1 & \text{if individual } i \text{ dies,} \\ 0 & \text{otherwise,} \end{cases}$$

for $i = 1, \ldots, L_{x,t}$. We define the time lived by individual i before $(x+1)th$ birthday by τ_i. We assume that we have at our disposal i.i.d. observations (δ_i, τ_i) for each of the $L_{x,t}$ individuals. Then,

$$\sum_{i=1}^{L_{x,t}} \tau_i = E_{x,t} \quad \text{and} \quad \sum_{i=1}^{L_{x,t}} \delta_i = D_{x,t}.$$

The probability of death at attained age x for the calendar year t, is denoted by $\widehat{q}_x(t)$ and computed according the Hoem estimator,

$$\widehat{q}_x(t) = \frac{D_{x,t}}{E_{x,t}}.$$

3   The methods

In the following, we present the methodological aspects of the four methods considered. This section relies on the technical report Tomas and Planchet (2013d).

The approach involving one parameter is the simplest methodology considered. It consists of applying a single factor of reduction / increase to the baseline probability of death denoted by $q_x^{\mathrm{b}}(t)$. In practice, this coefficient is the Standardized Mortality Ratio (SMR) of the population considered, see Liddell (1984), here computed on a global level. Then, we obtain the probabilities of death of the insurer, denoted by $\widetilde{q}_x(t)$, for $x \in [\underline{x}, \overline{x}]$ et $t \in [\underline{t}, \overline{t}]$ by

$$\widetilde{q}_x(t) = \mathrm{SMR} \times q_x^{\mathrm{b}}(t) \quad \text{with} \quad \mathrm{SMR} = \frac{\sum_{(x^*, t^*)} D_{x,t}}{\sum_{(x^*, t^*)} E_{x,t}\, q_x^{\mathrm{b}}(t)},$$

where x^* and t^* correspond to the age range and to the period of observation in common with the baseline mortality respectively.

In consequence, we adjust the mortality of the insurer only with one parameter, the SMR. It represents the observed deviation between the deaths recorded by the insurer and the ones obtained from the baseline mortality. The choice of x^* is of great importance because the table constructed is only valid on the age range considered.

3.1 Method 2: Approach involving two parameters with a semi-parametric relational model

The second approach is a semi-parametric Brass-type relational model. The fit is performed using the logistic function,

$$\text{logit } q_{x^*}(t^*) = \alpha + \beta \text{ logit } q_{x^*}^{\text{b}}(t^*).$$

where x^* and t^* correspond to the age range and to the period of observation in common with the baseline mortality respectively and $q_{x^*}^{\text{b}}(t^*)$ is the baseline mortality.

This model implies that the differences between the observed and baseline mortality can be represented linearly with two parameters. The parameter α is an indicator of mortality affecting all ages identically while the parameter β modifies this effect with age. The estimation is done by minimizing a weighted distance between the estimated and observed probabilities of death,

$$\sum \left| E_{x^*,t^*} \times \left(\widehat{q}_{x^*}(t^*) - \widetilde{q}_{x^*}(t^*) \right) \right|.$$

This model has the advantage of integrated estimation and forecasting, as the parameters α and β are constant. We refer to Planchet and Thérond (2011, Ch.7) for more details.

We obtain the probabilities of death of the insurer $\widetilde{q}_x(t)$, for $x \in \left[\underline{x}, \overline{x} \right]$ et $t \in \left[\underline{t}, \overline{t} \right]$ by

$$\widetilde{q}_x(t) = \frac{\exp\left(\widehat{\alpha} + \widehat{\beta} \text{ logit } q_x^{\text{b}}(t)\right)}{1 + \exp\left(\widehat{\alpha} + \widehat{\beta} \text{ logit } q_x^{\text{b}}(t)\right)}.$$

It is interesting to note that in the interval $\left]0, 1/2\right[$ the logit function is concave and by the Jensen inequality if a function $f(x)$ is concave then $\mathbb{E}\left[f(x)\right] \geq f\left(\mathbb{E}[x]\right)$. It implies that $\mathbb{E}\left[\text{logit } q_x(t,i)\right] \leq \text{logit } q_x(t,i)$. In consequence, the probabilities of death are under-estimated when $q_x(t,i) < 1/2$ and over-estimated reciprocally. Models using the logit of mortality data must be use with care in case of mortality risk as they under-estimated the mortality rates in sometimes large proportions.

3.2 Method 3: Poisson GLM including interactions with age and calendar year

The third approach is a Poisson Generalized Linear Model (GLM) including the baseline mortality as a covariate and allowing interactions with age and calendar year.
With the notation of Section 2, the likelihood becomes

$$\mathcal{L}\left(\widehat{q}_x(t)\right) = \exp\left(-E_{x,t} \, \widehat{q}_x(t)\right)\left(\widehat{q}_x(t)\right)^{D_{x,t}}.$$

The associated log-likelihood is

$$\ell\left(\widehat{q}_x(t)\right) = \log \mathcal{L}\left(\widehat{q}_x(t)\right) = -E_{x,t} \, \widehat{q}_x(t) + D_{x,t} \log \widehat{q}_x(t).$$

Maximizing the log-likelihood $\ell\left(\widehat{q}_x(t)\right)$ gives $\widetilde{q}_x(t) = D_{x,t}/E_{x,t}$ which coincides with the central death rates $\widehat{m}_x(t)$. Then it is apparent that the likelihood $\ell\left(\widehat{q}_x(t)\right)$ is proportional to the Poisson likelihood based on

$$D_{x,t} \sim \mathcal{P}\left(E_{x,t} \, \widehat{q}_x(t)\right). \tag{1}$$

Thus, it is equivalent to work on the basis of the *true* likelihood or on the basis of the Poisson likelihood, as recalled in Delwarde and Denuit (2005). In consequence, we consider (1) to take advantage of the GLMs framework.

We suppose then that the number of deaths of the insurer at attained age x^* and calendar year t^* is determined by

$$D_{x^*,t^*} \sim \mathcal{P}\left(E_{x^*,t^*}\,\mu_{x^*}(t^*)\right),$$

$$\text{with}\quad \mu_{x^*}(t^*) = \beta_0 + \beta_1 \log q_{x^*}^{\mathrm{b}}(t^*) + \beta_2\,x^* + \beta_3\,t^* + \beta_4\,x^*\,t^*,$$

where x^* and t^* correspond to the age range and to the period of observation in common with the baseline mortality respectively and $q_{x^*}^{\mathrm{b}}(t^*)$ are the baseline mortality.

If we do not allow for interactions, we will observe parallel shifts of the probabilities of death according to the baseline mortality for each dimension. This view is certainly unrealistic and interactions need to be incorporated. However, they can only be reasonably taken into account if we have at our disposal a large common historic with the baseline mortality.

We obtain the probabilities of death of the insurer $\widetilde{q}_x(t)$, for $x \in \left[\underline{x},\overline{x}\right]$ et $t \in \left[\underline{t},\overline{t}\right]$ by

$$\widetilde{q}_x(t) = \exp\left(\widehat{\beta}_0 + \widehat{\beta}_1 \log q_x^{\mathrm{b}}(t) + \widehat{\beta}_2\,x + \widehat{\beta}_3\,t + \widehat{\beta}_4\,x\,t\right).$$

It should be noticed that this method is not flexible when we forecast the mortality trends. In addition, it is also relatively unstable if we don't have a large historic at our disposal. This method should be used with care, especially when data have a high underlying heterogeneity.

3.3 Method 4: Non-parametric smoothing of the periodic table and application of the improvement rates of the baseline mortality

The fourth approach consists of, in a first step, smoothing the periodic table computed from the aggregated data of the insurer and, in a second step, the application of the rates of mortality improvement derived from the baseline mortality.

We consider the following non-parametric relational model applied to the periodic table of the insurer,

$$D_x \sim \mathcal{P}\left(E_x\,q_x^{\mathrm{b}}\,\exp(f(x))\right),$$

including the expected number of deaths $E_x\,q_x^{\mathrm{b}}$ according to the baseline mortality and where f is an unspecified smooth functions of attained age x.

Similarly to the Method 3, see Section 3.2, we are taking advantage of the GLMs framework. However, here, the role of the GLMs is of a background model which is fitted locally. We consider the local kernel-weighted log-likelihood method to estimate the smooth function $f(x)$. Statistical aspects of local likelihood techniques have been discussed extensively in Tomas (2013). These methods have been used in a mortality context by Delwarde *et al.* (2004), Debón *et al.* (2006) and Tomas (2011) to graduate life tables with attained age. More recently, Tomas and Planchet (2013a) have covered smoothing in two dimensions and introduced adaptive parameters choice with an application to long-term care insurance. Local likelihood techniques have the ability to model the mortality patterns even in presence of complex structures and avoid to rely on experts opinion.

The selection of the smoothing parameters, i.e. the window width λ and the polynomial degree p, is an effective compromise between two objectives, the elimination of irregularities and the achievement of a desired mathematical shape to the progression of the mortality rates. This underlines the importance of experience, and above all, of thorough investigation of data as the prerequisites of reliable judgment, as we must first inspect the data and take the decision as the type of irregularity we wish to retain. The strategy is to evaluate an number of candidates and to use a criteria to select among the fits the one with the lowest score. Here, we use the Akaike information criterion (AIC) based on the deviance.

In practice, the smoothing parameters are selected using graphical diagnostics. The graphic displays the AIC scores against the fitted degrees of freedom. This aids interpretation: 2 degrees of freedom represents a smooth model with very little flexibility while 10 degrees of freedom represents a noisy model showing many features. It also aids comparability as we can compute criteria scores for other polynomial degrees or for other smoothing methods and added them to the plot. We select the smoothing parameters at the point when the criterion reaches a minimum or a plateau after a steep descent.

Once the periodic table adjusted, the probabilities of death of the insurer $\widetilde{q}_x(t)$, for $x \in \left[\underline{x}, \overline{x}\right]$ et $t \in \left[\underline{t}, \overline{t}\right]$ are obtained by

$$\widetilde{q}_x(t) = q_x^{\mathrm{b}}(t) \, \exp\big(\widehat{f}(x)\big) \,.$$

This means that we apply the improvement rates, $q_x^{\mathrm{b}}(t)/q_x^{\mathrm{b}}$, derived from the baseline mortality to the periodic table $q_x^{\mathrm{b}} \exp\big(\widehat{f}(x)\big)$.

3.4   Completion of the tables, the approach of Denuit and Goderniaux (2005)

Finally, we need to complete the tables. Due to the probable lack of data beyond a certain age, we have not valid information to derive mortality at high ages. Actuaries and demographers have developed various techniques for the completion of the tables at high ages.

In this article, we use a simple and efficient method proposed by Denuit and Goderniaux (2005). This method relies on the fitted one year probabilities of death and introduces two constraints about the completion of the mortality table. It consists to fit, by ordinary least squares, the following log-quadratic model:

$$\log \widetilde{q}_x(t) = a_t + b_t \, x + c_t \, x^2 + \epsilon_x(t), \tag{2}$$

where $\epsilon_x(t) \sim$ iid Normal$(0, \sigma^2)$, separately for each calendar year t at attained ages $x^\star$. Two restrictives conditions are imposed:

i. Firstly, a completion constraint,

$$q_{130}(t) = 1 \,, \quad \text{for all } t.$$

Even though human lifetime does not seem to approach any fixed limit imposed by biological factors or other, it seems reasonable to accept the hypothesis that the age limit of end of life 130 will not be exceeded.

ii. Secondly, an inflexion constraint,

$$\frac{\partial}{\partial x} q_x(t)|_{x=130} = 0 \,, \quad \text{for all } t.$$

These constraints impose concavity at older ages in addition to the existence of a tangent at the point $x = 130$. They lead to the following relation between the parameters a_t, b_t and c_t for each calendar year t:

$$a_t + b_t \, x + c_t \, x^2 = c_t \, (130 - x)^2,$$

for $x = x_t^\star, x_t^\star + 1, \ldots$. The parameters c_t are estimated from the series $\{\widetilde{q}_x(t), x = x_t^\star, x_t^\star + 1, \ldots\}$ of calendar year t with equation (2) and the constraints imposed.

It should be noted that the completed parts of the table are rather formal and are not validated.

4 The validation

The validation is assessed on three levels. This concerns the proximity between the observations and the model, the regularity of the fit and the consistency and plausibility of the mortality trends. This section relies on the technical report Tomas and Planchet (2013c).

4.1 First level: proximity between the observations and the model

The first level of the validation assesses the overall deviation with the observed mortality. It involves the SMR test proposed by Liddell (1984), the likelihood ratio test and the Wilcoxon Matched-Pairs Signed-Ranks test.

In addition, we find useful to compare criteria measuring the distance between the observations and the models with the χ^2 applied by Forfar *et al.* (1988), the mean average percentage error (MAPE) and R^2 applied by Felipe *et al.* (2002) as well as the deviance, the SMR and the number of standardized residuals larger then 2 and 3.

The tests and quantities summarizing the overall deviation between the observations and the model are described in the following:

χ^2**.** This indicator allows to measure the quality of the fit of the model. It writes

$$\chi^2 = \sum_{(x,t)} \frac{\left(D_{x,t} - E_{x,t}\, \widetilde{q}_x(t)\right)^2}{E_{x,t}\, \widetilde{q}_x(t)\left(1 - \widetilde{q}_x(t)\right)}$$

We will privilege the model having the lowest χ^2.

MAPE. This is a measure of accuracy of the fit to the observations. This indicator is the average of the absolute values of the deviations from the observations,

$$\text{MAPE} = \frac{\sum_{(x,t)} \left|\left(D_{x,t}/E_{x,t} - \widetilde{q}_x(t)\right)/\left(D_{x,t}/E_{x,t}\right)\right|}{\sum_{(x,t)} D_{x,t}} \times 100.$$

It is a percentage and thus a practical indicator for the comparison. However, in presence of zero observations, there will be divisions by zero, and these observations must be removed.

R^2**.** The coefficient of determination measures the adequacy between the model and the observation. It is defined as the part of variance explained with respect to the total variance,

$$\text{R}^2 = 1 - \left(\frac{\sum_{(x,t)}\left(D_{x,t}/E_{x,t} - \widetilde{q}_x(t)\right)^2}{\sum_{(x,t)}\left(D_{x,t}/E_{x,t} - \left(\sum_{(x,t)}(D_{x,t}/E_{x,t})/n\right)\right)^2}\right),$$

where n is the number of observations.

The deviance. This is a measure of the quality of the fit. Under the hypothesis of the number of deaths following a Poisson law $D_{x,t} \sim \mathcal{P}\left(\text{E}_{x,t}\, q_x(t)\right)$, the deviance writes

$$\text{If } D_{x,t} > 0, \ \text{Deviance}_{x,t} = 2\left(D_{x,t} \ln\left(\frac{D_{x,t}}{E_{x,t}\, \widetilde{q}_x(t)}\right) - \left(D_{x,t} - E_{x,t}\, \widetilde{q}_x(t)\right)\right).$$

$$\text{If } D_{x,t} = 0, \ \text{Deviance}_{x,t} = 2\, E_{x,t}\, \widetilde{q}_x(t).$$

$$\text{And } \ \text{Global deviance} = \sum_{(x,t)} \text{Deviance}_{x,t}\,.$$

The likelihood ratio test. We can also consider the p-value associated to the likelihood ratio test (or drop-in deviance test).

We seek here to determine if the fit corresponds to the underlying mortality law (null hypothesis $\mathcal{H}_0$). The likelihood ratio test statistic, ξ^{LR}, writes

$$\xi^{\text{LR}} = \sum_{(x,t)} \left(D_{x,t} \ln \left(\frac{D_{x,t}}{E_{x,t} \, \widetilde{q}_x(t)} \right) - \left(D_{x,t} - E_{x,t} \, \widetilde{q}_x(t) \right) \right).$$

If $\mathcal{H}_0$ is true, this statistic follows a χ^2 law with a number of degrees of freedom equal to the number of observations n:

$$\xi^{\text{LR}} \sim \chi^2(n).$$

Hence, the null hypothesis $\mathcal{H}_0$ is rejected if

$$\xi^{\text{LR}} > \chi^2_{1-\alpha}(n),$$

where $\chi^2_{1-\alpha}(n)$ is the $(1-\alpha)$ quantile of the χ^2 distribution with n degrees of freedom. The p-value is the lowest value of the type I error (α) for which we reject the test. We will privilege the model having the closest p-value to 1,

$$\text{p-value} = \mathbb{P}\big[\chi^2_{1-\alpha}(n) > \xi^{\text{LR}}\big] = 1 - \text{F}_{\chi^2(n)}(\xi^{\text{LR}}).$$

The SMR. It is the ratio between the observed and fitted number of deaths. If we consider that the number of deaths follows a Poisson law $D_{x,t} \sim \mathcal{P}\big(E_{x,t} \, qx(t)\big)$,,

$$\text{SMR} = \frac{\sum_{(x,t)} D_{x,t}}{\sum_{(x,t)} E_{x,t} \, \widetilde{q}_x(t)}.$$

Hence, if SMR > 1, the fitted deaths are under-estimated and vice-versa if SMR < 1.

The SMR test. We can also apply a test to determine if the SMR is significatively different from 1, see Liddell (1984). We compute the following statistic,

$$\text{If SMR } > 1, \xi^{\text{SMR}} = 3 \times D^{\frac{1}{2}} \big(1 - (9D)^{-1} - (D/E)^{\frac{1}{3}} \big).$$
$$\text{If SMR } < 1, \xi^{\text{SMR}} = 3 \times D^{*\frac{1}{2}} \big((D^*/E)^{\frac{1}{3}} + (9D^*)^{-1} - 1 \big),$$

where $D = \sum_{(x,t)} D_{x,t}$, $D^* = \sum_{(x,t)} D_{x,t} + 1$ and $E = \sum_{(x,t)} E_{x,t} \, \widetilde{q}_x(t)$. If the SMR is not significatively different from 1 (null hypothesis $\mathcal{H}_0$), this statistic follows a standard Normal law,

$$\xi^{\text{SMR}} \sim \text{N}(0,1).$$

Thus, the null hypothesis $\mathcal{H}_0$ is rejected if

$$\xi^{\text{SMR}} > \text{N}_{1-\alpha}(0,1),$$

where $\text{N}_{1-\alpha}(0,1)$ is the $(1-\alpha)$ quantile of the standard Normal distribution. The p-value is given by

$$\text{p-value} = 1 - \text{F}_{\text{N}(0,1)}(\xi^{\text{SMR}}).$$

We will seek to obtain the closest p-value to 1.

The Wilcoxon Matched-Pairs Signed-Ranks test. The framework of this test is very similar to the signs

test, see Section .

While the signs test only uses the information on the direction of the differences between pairs composed by the observed and fitted probabilities of death, the Wilcoxon test takes in addition into account the magnitude of these differences. We test the null hypothesis $\mathcal{H}_0$ that the median between the difference of each pairs is null.

We compute the difference between the observed and fitted probabilities of death, and rank them by increasing order of the absolute values, omitting the null differences. We assigned to each non-zero difference its rank. We denote by w_+ et w_- the sum of the ranks of the differences strictly positive and negative respectively. Finally, we denote by w the maximum between the two numbers: $w = \max\{w_+, w_-\}$.

If the observed and fitted probabilities of death are equal, i.e. if $\mathcal{H}_0$ is true, the sum of the rank having a positive and negative sign should be approximatively equal. But if the sum of the ranks of positive signs differs largely from the sum of the ranks of the negative signs, we will deduce that the observed probabilities of death differ from the fitted ones, and we will reject the null hypothesis.

We can compute the statistic ξ^{WIL},

$$\xi^{\text{WIL}} = \frac{(w - 1/2 - n\,(n+1))/4}{\sqrt{n\,(n+1)\,(2\,n+1)/24}}.$$

If $\mathcal{H}_0$ is true, this statistic follows a standard Normal law,

$$\xi^{\text{WIL}} \sim \text{N}(0,1).$$

Thus the null hypothesis $\mathcal{H}_0$ is rejected if,

$$|\xi^{\text{WIL}}| > \text{N}_{1-\alpha/2}(0,1),$$

where $\text{N}_{1-\alpha/2}(0,1)$ is the $(1 - \alpha/2)$ quantile of the standard Normal distribution.

The p-value is given by:

$$\text{p-value} = \mathbb{P}\big[\text{N}_{1-\alpha/2}(0,1) > |\xi^{\text{WIL}}|\big] = 2 \times \big(1 - \text{F}_{\text{N}(0,1)}(|\xi^{\text{WIL}}|)\big).$$

We will seek to obtain the closest p-value to 1.

The Wilcoxon test uses the magnitude of the differences. The result may be different from the signs test that uses the number of positive and negative signs of the difference.

Besides the tests and quantities, the process of validation for the first level involves graphical analysis. It consists of representing graphically the fitted values against the observations for a given attained age or calendar year.

In conjunction with looking to the plots of the fits, we should study the residuals plots. We determine

Response residuals. $r_{x,t} = q_x(t) - \widetilde{q}_x(t)$;

Pearson residuals. $r_{x,t} = \big(D_{x,t} - E_{x,t}\,\widetilde{q}_x(t)\big)/\sqrt{\mathbb{V}\text{ar}\big[E_{x,t}\,\widetilde{q}_x(t)\big]}$;

Deviance residuals. $r_{x,t} = \mathrm{signe}\big(D_{x,t} - E_{x,t}\,\widetilde{q}_x(t)\big) \times \sqrt{\mathrm{Deviance}_{x,t}}$.

Such residual plots provide a powerful diagnostic that nicely complements the criteria. The diagnostic plots can show lack of fit locally and we have the opportunity to judge the lack of fit based on our knowledge of both the mechanism generating the data and of the performance of the model.

We superimposed a loess smooth curve on the response and Pearson residuals. These smooths help search for clusters of residuals that may indicate lack of fit. By reducing the noise, our attention may be more readily drawn to features that have been missed or not properly modeled by the approach. Here the process is not to judge a fit adequate if a smooth curve on its residuals plot is flat. A flat curve means simply that no systematic reproducible lack of fit has been detected. The fit may well be too noisy, and may stay too close to an interpolation. If the approach correctly models the data no strong patterns should appear in the response and Pearson residuals.

Moreover, if the Pearson residuals are mainly in the interval $[-2, 2]$, it indicates that the models adequately capture the variability of the dataset. If the deviance residuals exhibit several successive residuals having the same sign, it illustrates that the data have been over-smoothed locally.

Finally, a classical step of the validation consists of comparing the observed and fitted deaths for a given attained age or calendar year over the common period of observation. Using the usual Normal approximation of a Poisson law, we can confront graphically the observed and fitted deaths as well as the lower and upper bound of the pointwise confidence intervals.

Let suppose the following relation:

$$D_{x,t} \sim \mathcal{N}\big(E_{x,t}\,q_x(t), E_{x,t}\,q_x(t)\,(1 - q_x(t))\big).$$

An approximation of the $100\,\% \times (1 - \alpha)$ pointwise confidence intervals of $D_{x,t}$ is

$$\left(L_{x,t}\,\widetilde{q}_x(t) - z_{1-\alpha/2}\sqrt{E_{x,t}\,\widetilde{q}_x(t)\,\big(1 - \widetilde{q}_x(t)\big)},\ E_{x,t}\,\widetilde{q}_x(t) + z_{1-\alpha/2}\sqrt{E_{x,t}\,\widetilde{q}_x(t)\,\big(1 - \widetilde{q}_x(t)\big)} \right),$$

where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the Normal distribution.

The observed deaths must be inside the theoretical confidence intervals over the entire age range for the common period of observation, indicating a correct representation of the reality by the mortality table. We can also compute the relative error measured by the half-length on the confidence interval. This gives an indicator of the inaccuracy of the fit.

4.2 Second level: regularity of the fit

The second level of the validation assesses the regularity of the fit. It examines if the data have been over or under-smoothed. It involves the number of positive and negative signs of the response residuals and the corresponding signs test, the number of runs and the corresponding runs test used in Forfar *et al.* (1988) and Debón *et al.* (2006).

The tests and quantities summarizing the regularity of the fit are described in the following:

Signs test This is a non-parametric test that examines the frequencies of the signs changes of the difference between the observed and fitted probabilities of death. Under the null hypothesis $\mathcal{H}_0$, the median between the positive and negative signs of this difference is null. Let n_+ and n_- be the numbers of positive and negative signs respectively, with $n = n_+ + n_-$, the signs test statistic, ξ^{SIG}, writes:

$$\xi^{\mathrm{SIG}} = \frac{|n_+ - n_-| - 1}{\sqrt{n}}.$$

If $\mathcal{H}_0$ is true, this statistic follows a standard Normal law,

$$\xi^{\mathrm{SIG}} \sim \mathrm{N}(0,1).$$

Thus, the null hypothesis $\mathcal{H}_0$ is rejected if

$$|\xi^{\mathrm{SIG}}| > \mathrm{N}_{1-\alpha/2}(0,1),$$

where $\mathrm{N}_{1-\alpha/2}(0,1)$ is the $(1-\alpha/2)$ quantile of the standard Normal distribution. The p-value is given by

$$\text{p-value} = \mathbb{P}\big[\mathrm{N}_{1-\alpha/2}(0,1) > |\xi^{\mathrm{SIG}}|\big] = 2 \times \big(1 - \mathrm{F}_{\mathrm{N}(0,1)}(|\xi^{\mathrm{SIG}}|)\big).$$

We will seek to obtain the closest p-value to 1.

Runs test This is a non-parametric test which determines if the elements of a sequence are mutually independent. A run is the maximal non-empty segment of the sequence consisting of adjacent equal elements. For instance, the following sequence composed of 20 elements,

$$\{+ + - - - - + + + + - - + + + + - - - + +\},$$

is composed of 7 *runs* with 4 are composed of $+$ and 3 of $-$. Under the null hypothesis $\mathcal{H}_0$, the number of runs of a sequence of n elements is a random variable whose conditional distribution given the numbers n_+ and n_- of positive and negative signs, with $n = n_+ + n_-$ is approximatively Normal, with:

$$\mu = \frac{2\,n_+\,n_-}{n_+ + n_-} + 1 \qquad \text{and} \qquad \sigma^2 = \frac{2\,n_+\,n_-\,\big(2\,n_+\,n_- - (n_+ + n_-)\big)}{(n_+ + n_-)^2(n_+ + n_- - 1)}.$$

The runs test statistic, ξ^{RUN}, writes:

$$\xi^{\mathrm{RUN}} = \frac{\text{Number of runs} - \mu}{\sigma}.$$

If $\mathcal{H}_0$ is true, this statistic follows a standard Normal law,

$$\xi^{\mathrm{RUN}} \sim \mathrm{N}(0,1).$$

Hence, the null hypothesis $\mathcal{H}_0$ is rejected if

$$|\xi^{\mathrm{RUN}}| > \mathrm{N}_{1-\alpha/2}(0,1),$$

where $\mathrm{N}_{1-\alpha/2}(0,1)$ is the $(1-\alpha/2)$ quantile of the standard Normal distribution. The p-value is given by

$$\text{p-value} = \mathbb{P}\big[\mathrm{N}_{1-\alpha/2}(0,1) > |\xi^{\mathrm{RUN}}|\big] = 2 \times \big(1 - \mathrm{F}_{\mathrm{N}(0,1)}(|\xi^{\mathrm{RUN}}|)\big).$$

We will seek to obtain the closest p-value to 1. This test is also called the Wald-Wolfowitz test.

4.3 Third level: consistency and plausibility of the mortality trends

The third level of the validation covers the plausibility and consistency of the mortality trends. It is evaluated by singles indices summarizing the lifetime probability distribution for different cohorts at several ages such as cohort life expectancies $_{\omega}e_{\widetilde{x}}^{\nearrow}$, median age at death $\mathrm{Med}\big[_{\omega}T_{\widetilde{x}}\big]$ and the entropy $\mathrm{H}\big[_{\omega}T_{\widetilde{x}}\big]$.

It also involves graphical diagnostics assessing the consistency of the historical and forecasted periodic life expectancy $_\omega e_x^\uparrow(t)$.

In addition, if we have at our disposal the male and female mortality, we can compare the trends of improvement and judge the plausibility of the common evolution of mortality of the two genders.

We refer to the concept of *biological reasonableness* proposed by Cairns *et al.* (2006) as a mean to assess the coherence if the extrapolated mortality trends. We ask the question where the data are originating from and based on this knowledge, what mixture of biological factors, medical advances and environmental changes would have to happen to cause this particular set of forecasts.

It consists, initially, to obtain the survival function calculated from the completed tables, see Section 3.4, resulting from the different approaches. From the survival function, we can derive a series of markers summarizing the lifetime probability distribution. We are interested in the survival distribution of cohorts for a given age $\widetilde{x}$ at time t. Hence, we are working along the diagonal of the Lexis diagram. As a result, we can determine the mortality trends and compare the level and the speed of improvement between the models.

We expose the indices summarizing the lifetime probability in the following:

Survival function. The survival function of a cohort aged $\widetilde{x}$ at time t measure the proportion of individuals of the cohort aged $\widetilde{x}$ at time t being alive at age $\widetilde{x} + x$ (or equivalently at time $t + x$). Under the condition $S_{\widetilde{x}}^{\nearrow}(0) = 1$, it writes,

$$S_{\widetilde{x}}^{\nearrow}(x) = \prod_{j=0}^{x-1} \big(1 - q_{\widetilde{x}+j}(t+j)\big) \, .$$

where the upper indices $\nearrow$ recalls that we are working along a diagonal of the Lexis diagram.

Cohort life expectancy. It is the partial life expectancy (over ω years) of an individual of a cohort aged $\widetilde{x}$ at time t. It writes,

$$_\omega e_{\widetilde{x}}^{\nearrow} = \int_1^\omega S_{\widetilde{x}}^{\nearrow}(u) \, du \, .$$

We obtain,

$$_\omega e_{\widetilde{x}}^{\nearrow} = \sum_{x=1}^{\omega} \prod_{j=0}^{x-1} \big(1 - q_{\widetilde{x}+j}(t+j)\big) \, .$$

Periodic life expectancy. It is the residual life expectancy (over ω years) of an indivudal aged x at time t. It writes,

$$_\omega e_x^\uparrow(t) = \sum_{\delta=1}^{\omega} \prod_{j=0}^{\delta-1} \big(1 - q_{x+j}(t)\big) \, ,$$

where the upper indices $\uparrow$ recalls that we are working along a vertical of the Lexis diagram.

Median age at death. The median age at death of an individual of a cohort aged $\widetilde{x}$ at time t (over ω years), denoted $\mathrm{Med}\big[_\omega T_{\widetilde{x}}\big]$, is the median of the lifetime probability distribution $T_{\widetilde{x}}$,

$$S_{\widetilde{x}}^{\nearrow}\big(\mathrm{Med}\big[_\omega T_{\widetilde{x}}\big]\big) = 0.5 \, .$$

Entropy. The entropy, $\mathrm{H}\big[{}_\omega T_{\widetilde{x}}\big]$, is the mean of $\ln S_{\widetilde{x}}^{\nearrow}(x)$ weighted by $S_{\widetilde{x}}^{\nearrow}(x)$ (over ω years),

$$\mathrm{H}\big[{}_\omega T_{\widetilde{x}}\big] = -\frac{\int_1^\omega S_{\widetilde{x}}^{\nearrow}(u)\,\ln S_{\widetilde{x}}^{\nearrow}(u)\,du}{\int_1^\omega S_{\widetilde{x}}^{\nearrow}(u)\,du}\,.$$

When the deaths become more concentrated, $\mathrm{H}\big[{}_\omega T_{\widetilde{x}}\big]$ decreases. In particular, $\mathrm{H}\big[{}_\omega T_{\widetilde{x}}\big] = 0$ if the survival function has a perfect rectangular shape.

5 Application

The four methods of adjustment as well as the three levels of the process of validation are illustrated in the following. Table 12 and Figures 14 et 15, in Appendix A, present the observed characteristics of the male and female population of the portfolio.

5.1 The methods of adjustment

5.1.1 Method 1

The first approach is the simplest methodology considered. It consists of applying the SMR to the probability of death of the baseline mortality. Table 1 presents the values of the SMR computed on a global level for different age ranges. For our application, we used the age range $\big[30, 95\big]$.

	Age range				
	$\big[30, 95\big]$	$\big[35, 90\big]$	$\big[40, 85\big]$	$\big[45, 80\big]$	$\big[60, 80\big]$
Male pop.	1.1830	1.1863	1.1898	1.2000	1.1181
Female pop.	1.1053	1.1068	1.1131	1.1281	1.0441

Table 1: Values of the SMR for different age ranges.

5.1.2 Method 2

The second method is a semi-parametric Brass-type relational model, implying that the differences between the observed and baseline mortality can be represented linearly with two parameters. Table 2 presents the values of the parameters α et β of the logistic model for the male and female population.

	α	β
Male pop.	-0.6455	0.8399
Female pop.	-0.4087	0.9115

Table 2: Parameters of the logistic model.

5.1.3 Method 3

The third approach is a Poisson GLM including the baseline mortality as a covariate and allowing interactions with age and calendar year.

For our illustration, we only have at our disposal three calendar years in common with the baseline mortality. In consequence, the interactions with the calendar year are not considered.

Tables 3 et 4 present the values of the parameters β_0, β_1 et β_2 of the Poisson GLM for the male and female population respectively.

Coef.	Par. est.	Err. Std.	z value	p-value
β_0	-3.5791	2.3162	-1.545	0.1223
β_1	0.5959	0.2023	2.946	0.0032
β_2	0.0271	0.0206	1.316	0.1882

Table 3: Results of the Poisson GLM, male population.

Coef.	Par. est.	Err. Std.	z value	p-value
β_0	4.6759	1.5324	3.051	0.0023
β_1	1.3259	0.1236	10.729	$< 2e - 16$
β_2	-0.0432	0.0133	-3.247	0.0012

Table 4: Results of the Poisson GLM, female population.

5.1.4 Method 4

The fourth approach includes in a first step a non-parametric smoothing of the periodic table derived from the aggregated data of the portfolio.

We consider the local kernel-weighted log-likelihood smoothing method. The smoothing parameters are selected with aid of graphical diagnostics.

Figures 2a and 2b present the values of the AIC criterion according the fitted degrees of freedom υ_2.
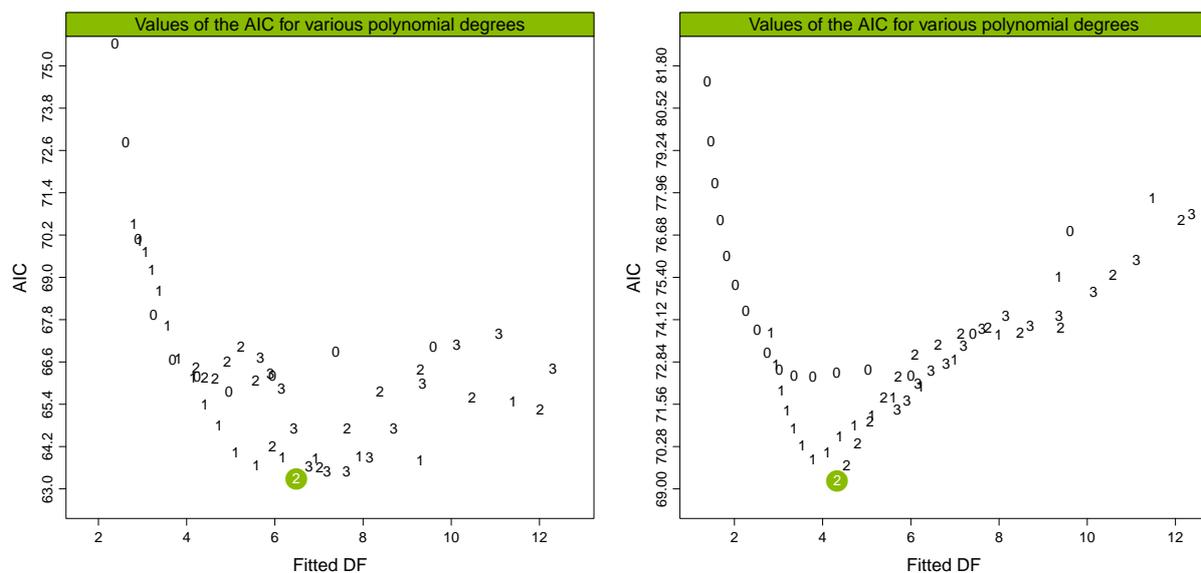
(a) Male population.

(b) Female population.

Figure 2: Values of the AIC criterion according the fitted degrees of freedom υ_2.

We choose the parameters when the AIC criterion reaches a minimum or a plateau after a steep descent. We select a quadratic fit with $v_2 = 6.48$ and 4.33, corresponding to a window width of 27 and 41 observations for the male and female population respectively.

Table 5 displays the chosen smoothing parameters.

	AIC	Fitted DF	λ	p
Male pop.	63.28	6.48	27	2
Female pop.	69.24	4.33	41	2

Table 5: Parameters of the local likelihood for smoothing the periodic tables.

Once the periodic tables adjusted, we apply the improvement rates derived from the baseline mortality to obtain the extrapolated probabilities of death.

5.1.5 Completion method

We adjust the quadratic constraint regression (2). The optimal starting age x^* is selected over the range $[75, 85]$. It turns out to be 85 for the four models.

The R^2 and corresponding estimated regression parameters c_t for the male and female population are displayed in Figure 3 for the four methods respectively.

The models capture more than 99.9 % of the variance of the probabilities of death at high ages, see Figures 3a, 3b, 3c and 3d top panel, for both populations. The regression parameters $\hat{c}_t$ decrease relatively linearly with the calendar year, see Figures 3a, 3b, 3c and 3d bottom panel.
This indicator represents the evolution of the mortality trends at high ages. We see that the mortality at high ages decreases and the speed of improvement is relatively similar for both populations except for the third method where the improvement is faster for the female than the male population.

We keep the original $\widetilde{q}_x(t)$ for ages below 85 years old for both populations, and replace the annual probabilities of death beyond this age by the values obtained from the quadratic regression. Results for the calendar year 2007 are presented in Figures 3e, 3f, 3g and 3h for both populations.

5.2 The process of validation

5.2.1 First level

Figures 4 and 5 present the fitted probabilities of death in original and log scale obtained by the four methods for the age range $[30, 95]$ and calendar year 2007 for both populations. These graphics give a first indication about the quality of the fits.
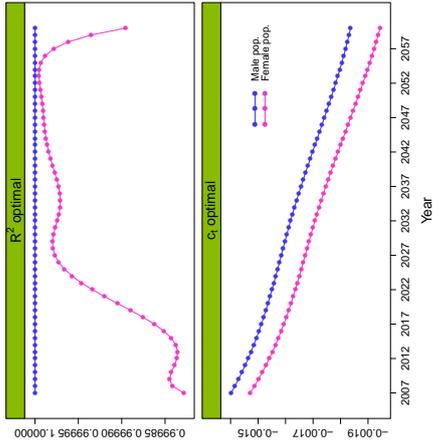
(a) R^2 et estimated c_t, method 1.
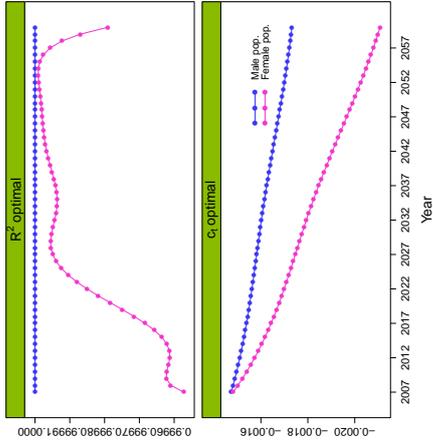
(b) R^2 et estimated c_t, method 2.
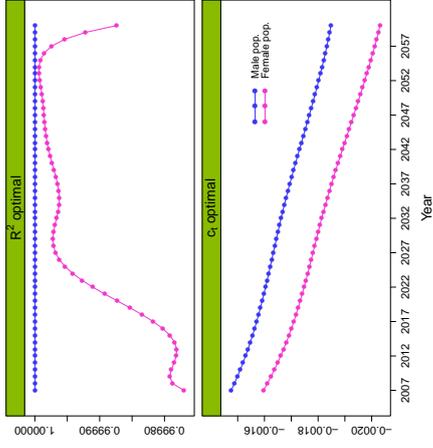
(c) R^2 et estimated c_t, method 3.
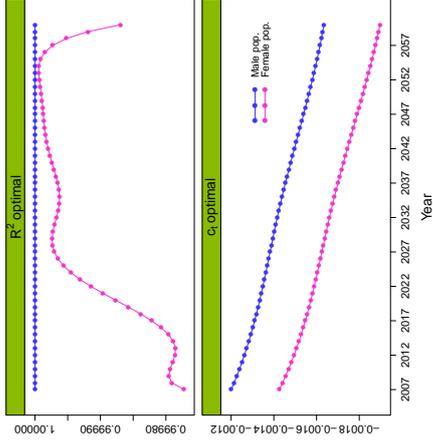
(d) R^2 et estimated c_t, method 4.
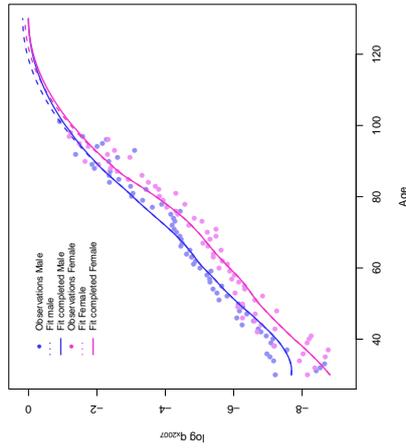
(e) Fits before and after the completion, method 1.
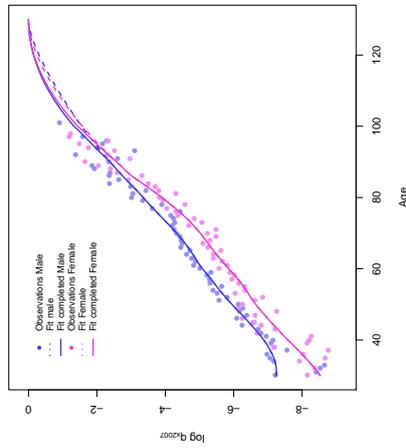
(f) Fits before and after the completion, method 2.
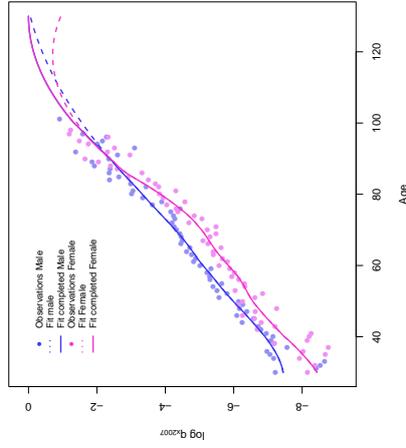
(g) Fits before and after the completion, method 3.
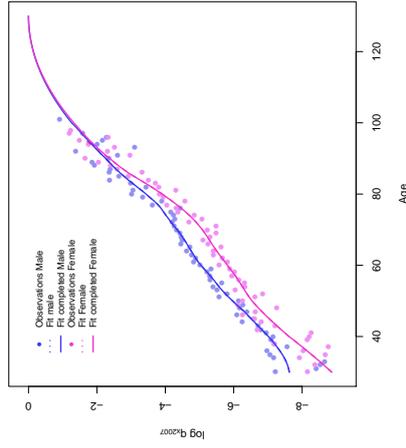
(h) Fits before and after the completion, method 4.

Figure 3: Regression parameters and fits obtained before and after the completion for the calendar year 2007 with the four methods.

(a) Male population

(b) Female population
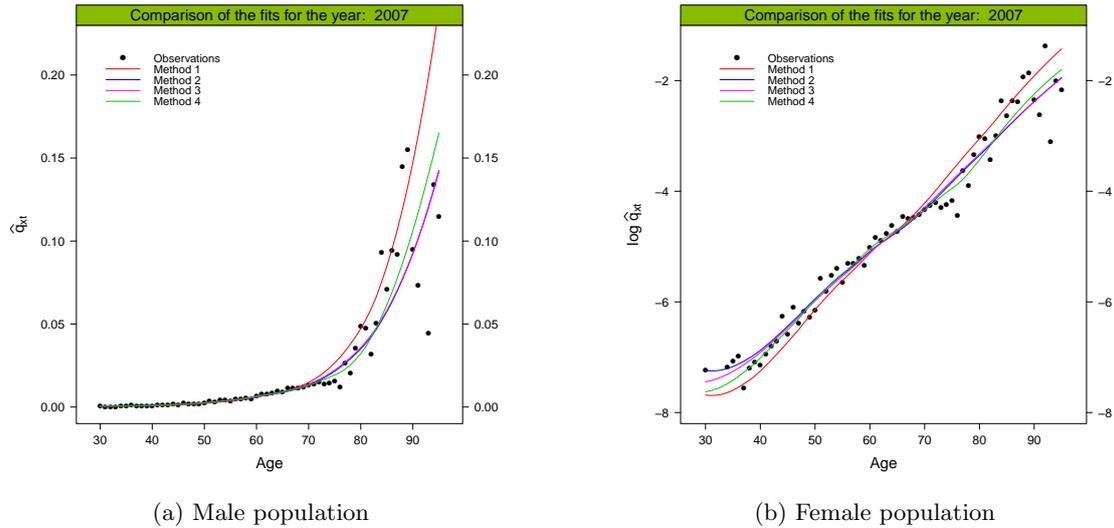
Figure 4: Fitted probabilities for calendar year 2007.

(a) Male population

(b) Female population
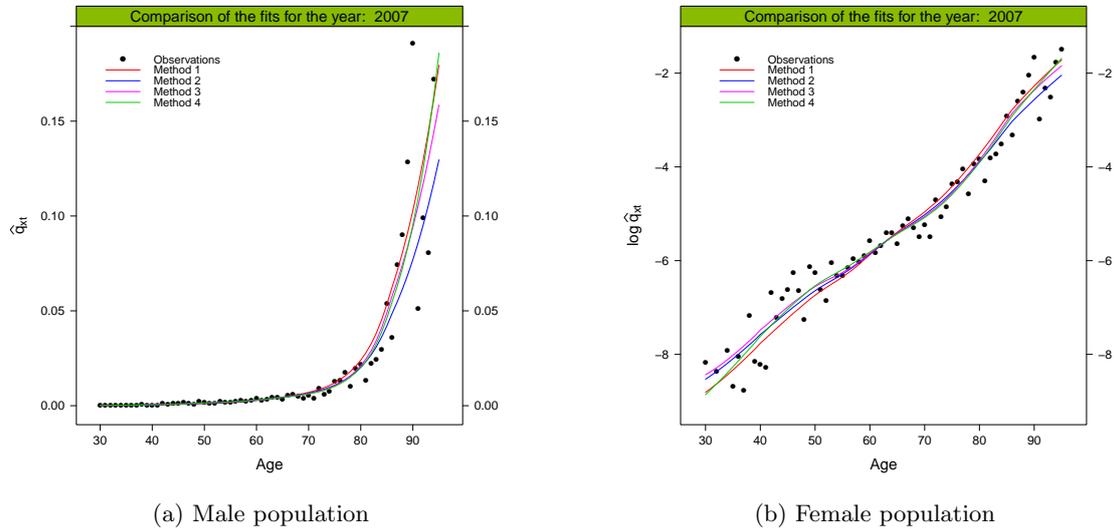
Figure 5: Fitted probabilities for calendar year 2007, log scale.

The plots of the residuals obtained by the four methods for calendar year 2007 are displayed in Figures 6a and 6b, for the male and female population respectively. We can visualize the quality of the fits obtained by the four methods for all the years in common with the baseline mortality and for both populations in Figures 7a to 7f.

For the male population, the Pearson and response residuals for the four methods display a slightly marked trend on the age range $[75, 90]$ and $[35, 50]$ than the other ages for the male and female population respectively. It indicates an inappropriate fit in this region. In addition, the Pearson residuals obtained with method 1 present a high curvature over the entire age range for both populations. We can visualize this lack of fit in the deviance residuals for the male population for the methods 2, 3 and 4 where several successive residuals exhibit a negative sign followed by a positive one for the age ranges $[71, 77]$ and $[81, 89]$ respectively. It indicates that the models over-estimate the probabilities of death for the age range $[71, 77]$ et under-estimate them for the age range $[81, 89]$.
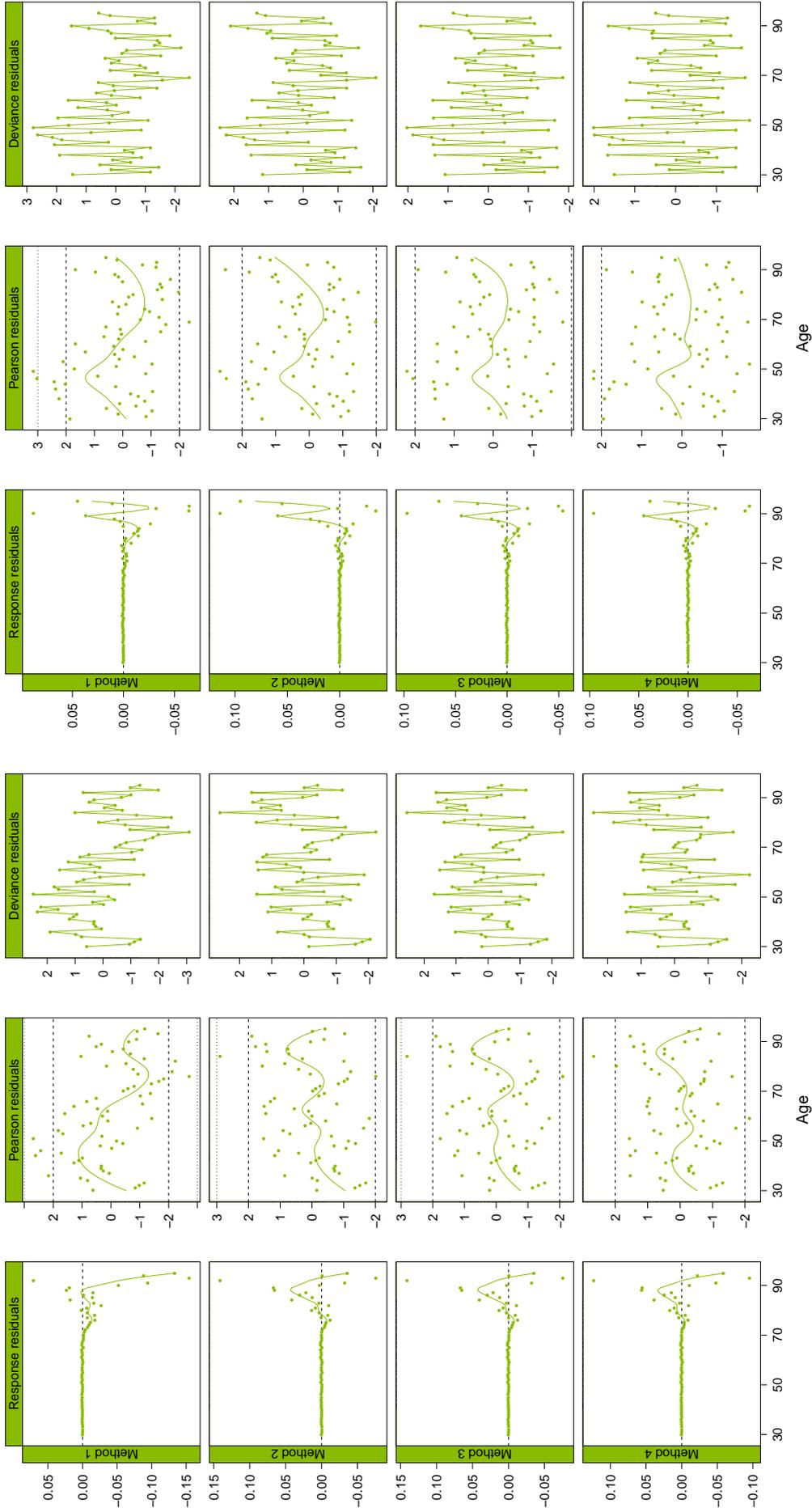
(a) Male population

(b) Female population

Figure 6: Comparison of the residuals for calendar year 2007.
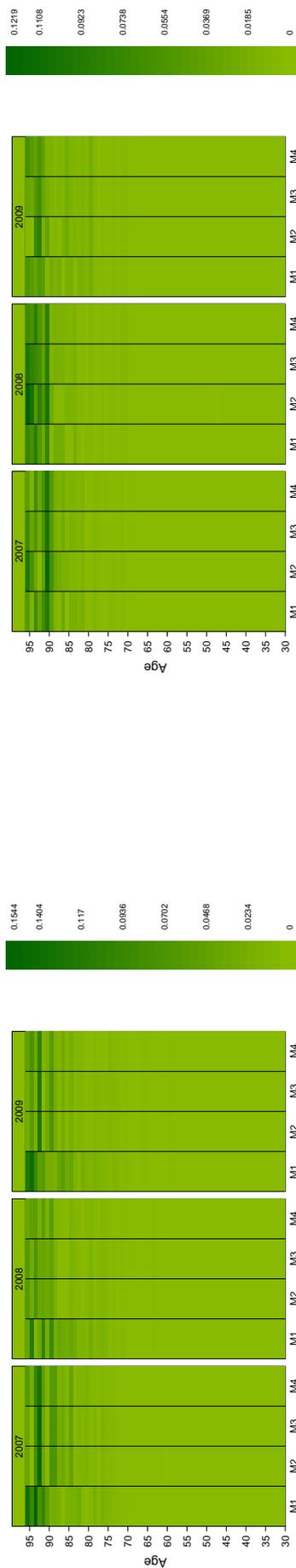
(a) Response residuals, male population.

(b) Response residuals, female population.
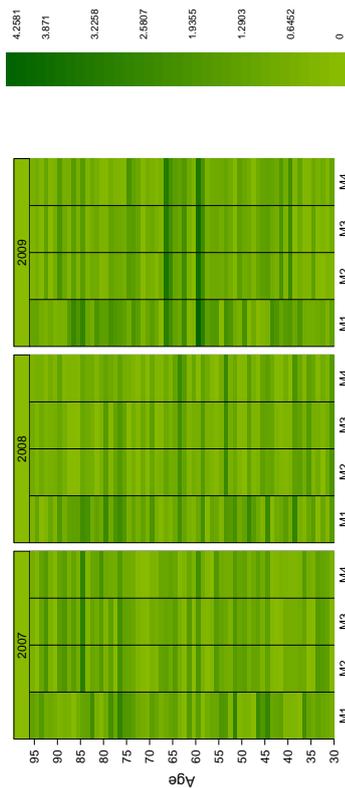
(c) Pearson residuals, male population.
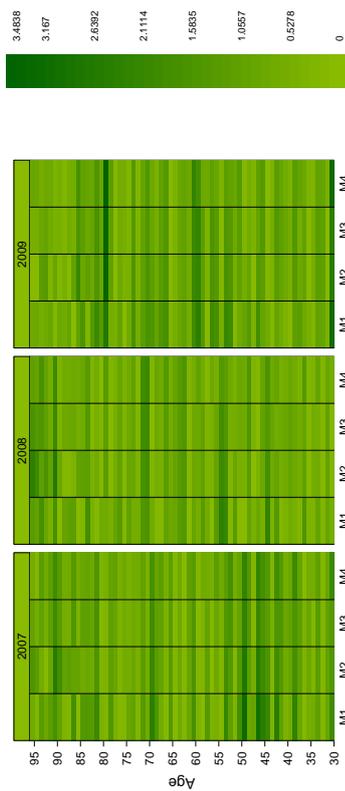
(d) Pearson residuals, female population.

(e) Deviance residuals, male population.
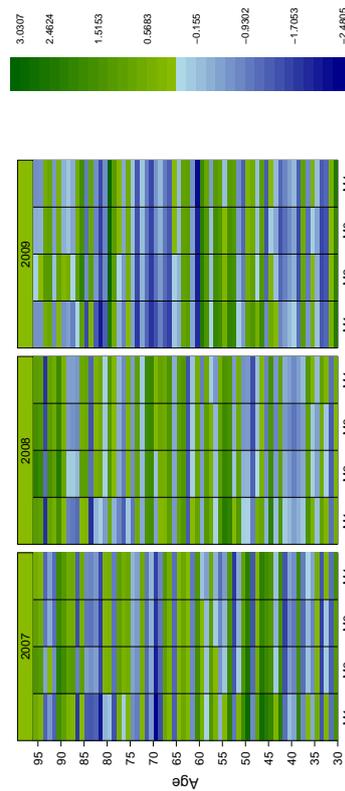
(f) Deviance residuals, female population.

Figure 7: Comparison of the residuals for calendar years 2007 to 2009.

For model 1, we observe several successive residuals with a positive and negative sign for large age ranges $[34, 47]$ and $[58, 80]$ showing a clear lack of fit. For the female population, methods 3 and 4 exhibit deviance residuals that articulate relatively correctly around 0. In addition, the Pearson residuals are principally in the interval $[-2, 2]$, showing that the models capture adequately the variability of the data.

We compare the fitted probabilities of death to the observed ones by carrying out the proposed tests and quantities presented in Section 4.1. Tables 6 and 7 present the results for the male and female population respectively.

		Method 1	Method 2	Method 3	Method 4
Standardized residuals	> 2	28	10	10	7
	> 3	1	2	2	2
χ^2		354.73	215.07	212.30	193.50
R^2		0.5573	0.7861	0.7873	0.7952
MAPE (%)		38.06	29.61	28.49	26.41
Deviance		369.30	231.680	225.76	202.08
Likelihood ratio test	ξ^{LR}	356.77	235.95	218.15	194.17
	p-value	0	0.0336	0.1554	0.5636
SMR test	SMR	1	1.0021	1	0.9998
	ξ^{SMR}	0.0044	0.1591	0.0088	0.0086
	p-value	0.4983	0.4368	0.4965	0.4966
Wilcoxon test	w	13284	10849	11173	10584
	ξ^{WIL}	4.2523	1.2362	1.6375	0.9079
	p-value	0	0.2291	0.1015	0.3639

Table 6: First level of validation, male population.

In a general manner, the four methods gives acceptable results. However, we remark that the quality of the fit increases according the complexity of the models. The non-parametric method 4, being the more flexible, leads to the most satisfying results. The model presents, for both populations, the lowest deviance, $MAPE$ and χ^2, as well as the lowest number of standardized residuals lower than 2. In addition, it exhibits the highest p-value for the Wilcoxon test and likelihood ratio test.

Figures 8 et 9 compare the observed and fitted deaths for the age range $[30, 95]$ and the years $[2007, 2009]$ for the male and female population respectively. The pointwise confidence intervals at 95% of the fitted deaths have been added to the plots.

The observed deaths are mainly within the bands of the theoretical pointwise confidence intervals at 95% over the age range considered. This indicates a correct representation of the reality by the male and female mortality tables. However, method 1 under-estimates the deaths for the age range $[30, 60]$ and strongly over-estimates them after 60 years old for both populations.

		Method 1	Method 2	Method 3	Method 4
Standardized residuals	> 2	15	9	5	5
	> 3	2	1	1	2
χ^2		250.40	226.44	199.00	196.60
R^2		0.8571	0.7750	0.8414	0.8568
MAPE (%)		31.63	30.86	33.00	30.70
Deviance		249.01	226.82	213.59	198.98
Likelihood ratio test	ξ^{LR}	240.01	300.01	203.19	188.74
	p-value	0.0222	0	0.3851	0.6697
SMR test	SMR	1	1.0287	1	0.9994
	ξ^{SMR}	0.0061	1.5410	0.0121	0.0210
	p-value	0.4976	0.0617	0.4952	0.4916
Wilcoxon test	w	10542	10767	9994	9950
	ξ^{WIL}	0.8559	1.1346	0.1771	0.1226
	p-value	0.4973	0.2565	0.8594	0.9024

Table 7: First level of validation, female population.

(a) Year 2007　　　　　(b) Year 2008　　　　　(c) Year 2009
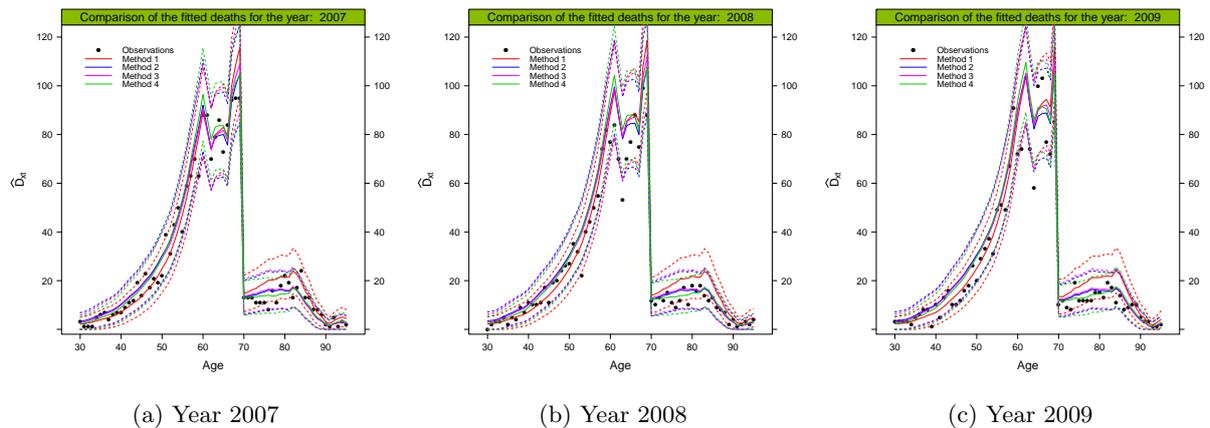
Figure 8: Comparison of the observed and fitted deaths, male population.

(a) Year 2007　　　　　(b) Year 2008　　　　　(c) Year 2009
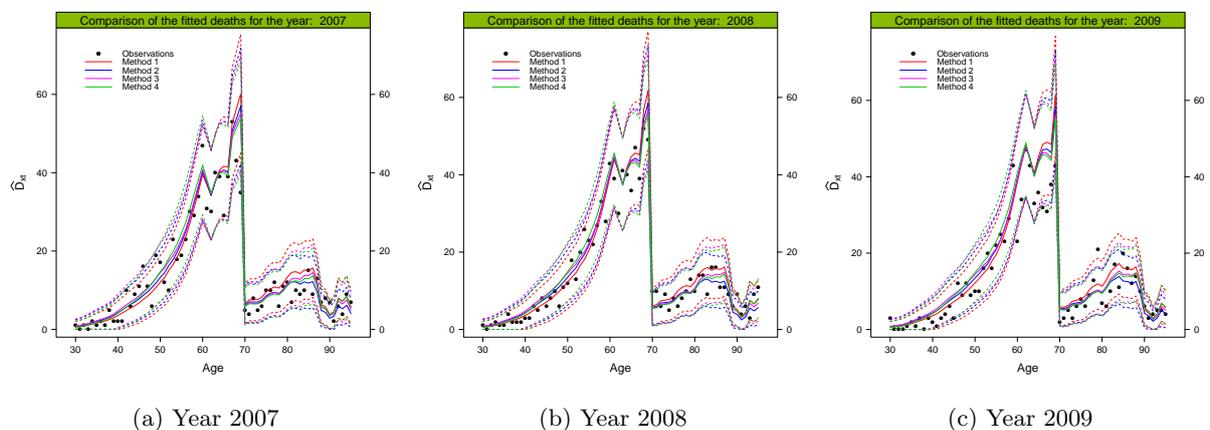
Figure 9: Comparison of the observed and fitted deaths, female population.

Table 8 presents relative error over the entire age range. We observe that the inaccuracy is important, being around 45 % and 31 % for the male and female population respectively.

	Method 1	Method 2	Method 3	Method 4
Male pop.	46.09	45.56	45.12	44.27
Female pop.	30.93	30.73	32.95	31.05

Table 8: Average relative error on the observed deaths (in %).

5.2.2 Second level

We compare the regularity of the fits obtained by the methods by carrying out the proposed tests and quantities presented in Section 4.2. Tables 9 and 10 present the results for the male and female population respectively.

		Method 1	Method 2	Method 3	Method 4
Signs test	$+(-)$	92(106)	86(112)	87(111)	96(102)
	ξ^{SIG}	0.9239	1.7767	1.6345	0.3553
	p-value	0.3556	0.0756	0.1021	0.7223
Runs test	Nber of $runs$	60	89	86	90
	$|\xi^{\mathrm{RUN}}|$	5.6577	1.3476	1.8145	1.4133
	p-value	0	0.1778	0.0696	0.1576

Table 9: Second level of validation, male population.

		Method 1	Method 2	Method 3	Method 4
Signs test	$+(-)$	97(101)	97(101)	93(105)	95(103)
	ξ^{SIG}	0.2132	0.2132	0.7817	0.4975
	p-value	0.8312	0.8312	0.4344	0.6189
Runs test	Nber of runs	94	102	96	104
	$|\xi^{\mathrm{RUN}}|$	0.8496	0.2132	0.502	0.5940
	p-value	0.3956	0.8312	0.6030	0.5525

Table 10: Second level of validation, female population.

All the methods gives raisonable results. However, only methods 2, 3 and 4 pass this step of the validation, having p-values greater than 5 %. The non-parametric method 4, having the capability of showing more features than the other models, leads to the highest number of runs.

5.2.3 Third level

From the survival function, we derive a series of markers summarizing the lifetime probability distribution. We are interested in the survival distribution of cohorts aged $\widetilde{x} = 30$ to 80 years old in 2007 over 50 years.

Table 11 presents the cohorts life expectancies, the median age at death and the entropy for both populations. Methods 2, 3 and 4 give relatively similar results. Conversely, the indices summarizing the lifetime probability distribution obtained by the method 1 stand out.

In a general manner, for both populations, method 1 leads to the smallest life expectancies for cohorts aged 50 and more in 2007. In consequence, the median age at death is the lowest and the deaths after 60 years old are the most stretched. Method 2 produces the highest life expectancies for high aged cohorts for both populations. It leads then to the highest median age at death and the most concentrated deaths after 80 years old.

Figures 10 and 11 compare the trends in periodic life expectancies for the ages 60, 75 and 80 for the male and female population respectively.

(a) Age 55. (b) Age 70. (c) Age 85.
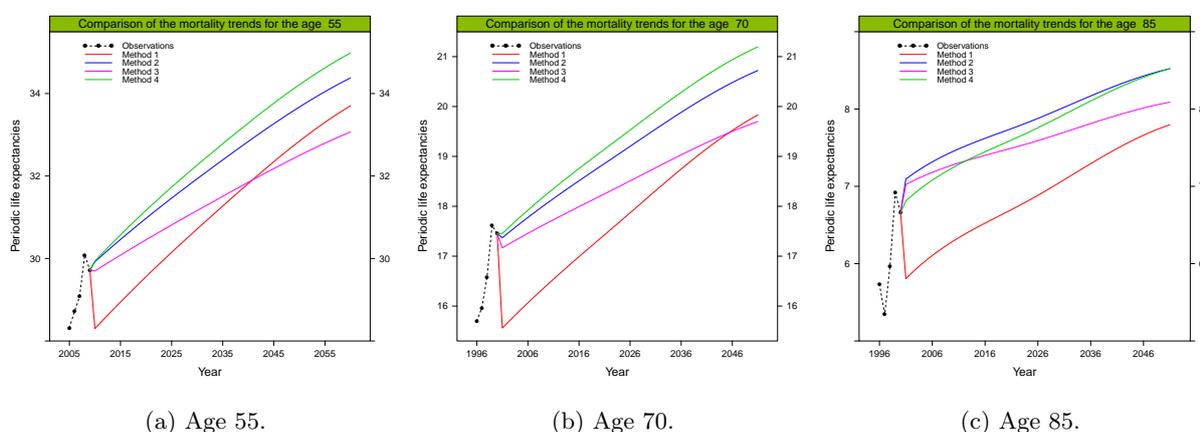
Figure 10: Comparison of the trends in periodic life expectancies by age for the male population.

(a) Age 55. (b) Age 70. (c) Age 85.
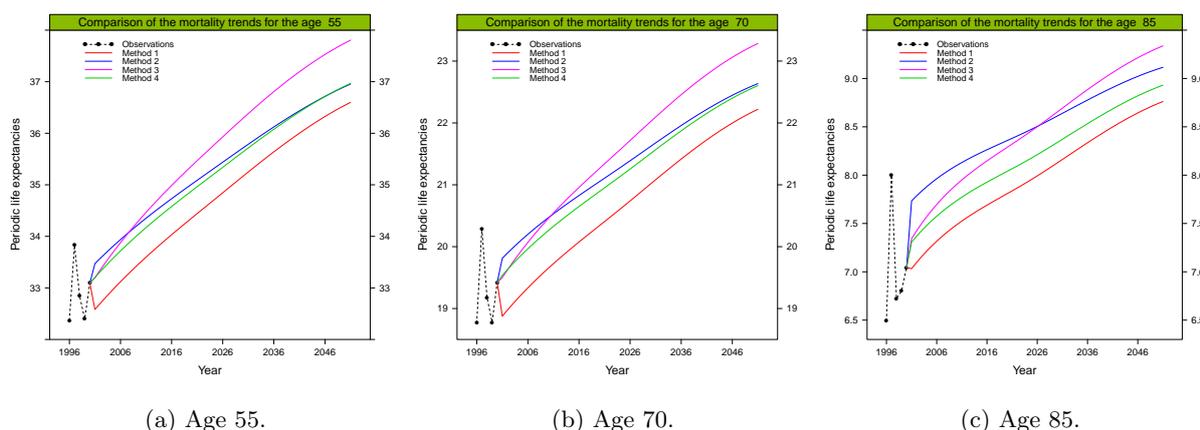
Figure 11: Comparison of the trends in periodic life expectancies by age for the female population.

The predicted periodic life expectancies obtained by methods 2, 3 and 4 follow adequately the observed trend. Only, the ones derived from method 1 seems to deviate. Moreover, methods 2 and 4 give relatively similar periodic life expectancies.

	Male population				Female population			
	Method 1	Method 2	Method 3	Method 4	Method 1	Method 2	Method 3	Method 4
$_{50}e_{30}\nearrow$	47.71	47.35	47.09	47.63	48.82	48.72	48.89	48.77
$_{50}e_{40}\nearrow$	44.07	44.23	43.58	44.64	46.79	46.89	47.32	46.92
$_{50}e_{50}\nearrow$	36.05	37.58	36.62	37.81	40.69	41.53	42.16	41.22
$_{50}e_{60}\nearrow$	25.89	28.13	27.32	28.13	30.69	32.07	32.34	31.48
$_{50}e_{70}\nearrow$	16.72	19.22	18.68	19.17	20.83	22.34	22.22	21.65
$_{50}e_{80}\nearrow$	9.17	11.71	11.43	11.22	11.90	13.45	12.88	12.52
Med$\left[_{50}T_{50}\right]$	39.88	42.52	41.09	42.66	44.77	46.36	46.93	45.52
Med$\left[_{50}T_{60}\right]$	28.75	31.63	30.60	31.65	33.71	35.44	35.57	34.54
Med$\left[_{50}T_{70}\right]$	18.61	21.49	21.04	21.57	23.15	24.98	24.70	24.01
Med$\left[_{50}T_{80}\right]$	10.18	13.19	12.88	12.56	13.32	15.17	14.44	14.03
H$\left[_{50}T_{30}\right]$	0.0009	0.0011	0.0012	0.0010	0.0005	0.0005	0.0004	0.0005
H$\left[_{50}T_{40}\right]$	0.0028	0.0027	0.0031	0.0025	0.0013	0.0013	0.0011	0.0013
H$\left[_{50}T_{50}\right]$	0.0111	0.0084	0.0098	0.0083	0.0056	0.0047	0.0042	0.0051
H$\left[_{50}T_{60}\right]$	0.0490	0.0338	0.0382	0.0350	0.0278	0.0222	0.0215	0.0251
H$\left[_{50}T_{70}\right]$	0.2179	0.1478	0.1618	0.1552	0.1345	0.1085	0.1096	0.1228
H$\left[_{50}T_{80}\right]$	1.0274	0.6833	0.7243	0.7417	0.6832	0.5495	0.5814	0.6302

Table 11: Single indices summarizing the lifetime probability distribution for cohorts of several ages in 2007 over 50 years.

Figures 12a to 12h display the ratio between the observed and predicted periodic life expectancies for both populations and for the four methods.

The observed periodic life expectancies are in a general manner higher than the ones predicted by the four methods for both populations before 70 years old and conversely. The approaches more under-estimated the periodic life expectancies of the male than the female population. In addition, the gap increases with the age. It reaches 5 % at 80 years old and 10 % to 15 % at 90 years old.

In the following, we compare the mortality trends of the male and female population to judge the plausibility of the common improvement. Figures 13a to 13d display the ratio between the cohorts life expectancies over 5 years of the two genders for the four methods. Figures 13e to 13h present the ratio between the fitted probabilities of death of the two genders.

The cohorts life expectancies of the female population are 15 % larger than the male ones at 110 years old, 10 % at 100 and 5 % at 90 years old, see Figures 13a, 13b and 13d. Methods 1, 2 and 4 lead to relatively similar mortality trends. The ratio tends to get closer to 1 with the calendar year, indicating that the male mortality is improving more rapidly than the female, which seems to be coherent, conversely with the trends obtained at high ages with method 3. The ratio between the female and male cohorts life expectancies obtained by the method 3 tends to move away from 1 with the calendar year at all ages, Figure 13c, indicating that the female mortality is improving more rapidly than the male. These remarks can be seen in Figures 13a to 13b as well. Moreover, none of the methods induces a higher female mortality than the male.

6   Conclusion

We have presented and illustrated the construction and validation of projected mortality table based on the specific information provided by the insurer to take into account his own mortality experience for the computation of his best estimate reserves.

Four methods of increasing complexity have been considered. We must acknowledge that our approaches may not be satisfactory enough in every situations. However, it is ensured with the fourth method of a proper representation of the periodic mortality, then we are only assuming that the rates of mortality improvement would follow the ones derived from the baseline mortality.

The validation have been assessed on three levels. The two first levels of the validation evaluate the fit according its regularity and the overall deviation from the past mortality. A satisfying fit, characterized by an homogeneous repartition of positive and negative signs of the response residuals and a high number of runs, should not result in a significant gap with the past mortality, or vice versa. Accordingly, the two first levels of the validation balance these two complementary aspects. The third level covers the plausibility and consistency of the mortality trends.

In addition, a practical guide, implementing the methodologies, based on the software R, R Development Core Team (2014) and on the package ELT, see Tomas and Planchet (2014), is available at www.ressources-actuarielles.net/gtmortalite.

This operational approach has been adopted by Institut des Actuaires as a reference framework (best practice) for organizations that need to construct entity specific projected mortality tables.
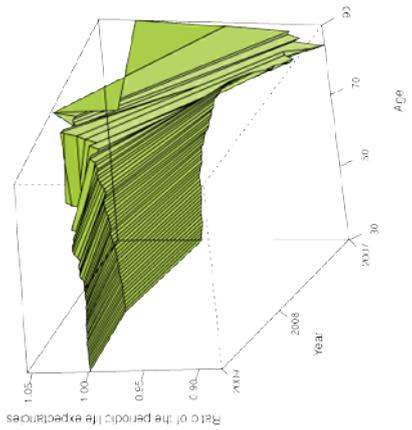
(a) Method 1, male population.
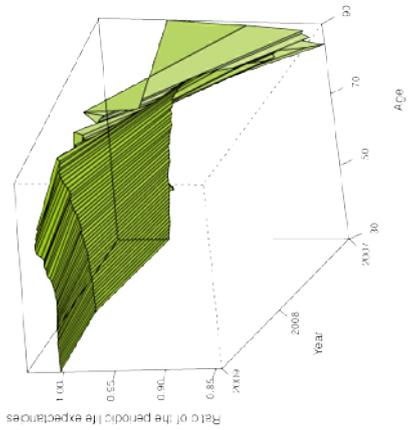
(b) Method 2, male population.
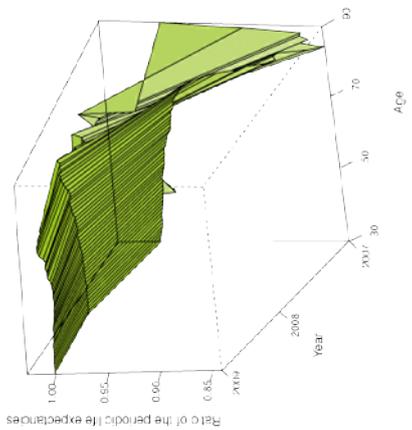
(c) Method 3, male population.
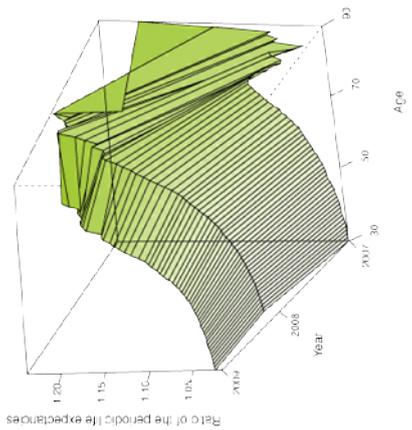
(d) Method 4, male population.
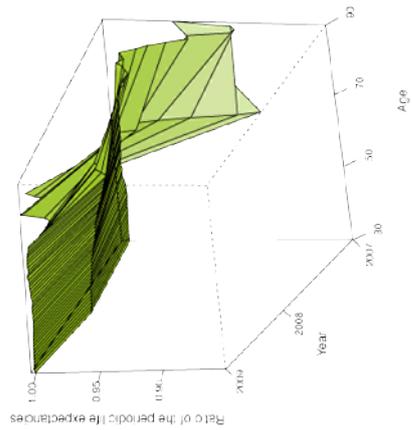
(e) Method 1, female population.
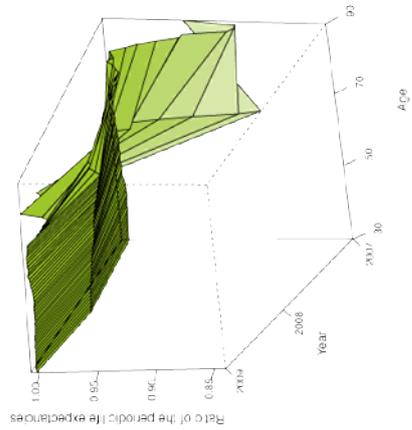
(f) Method 2, female population.
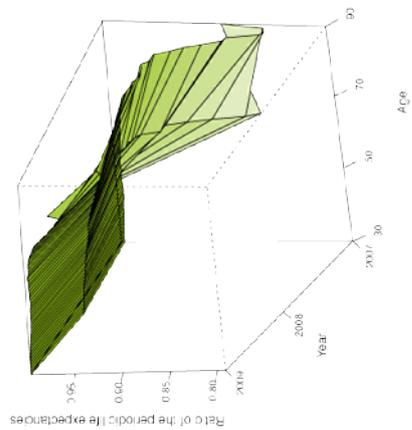
(g) Method 3, female population.
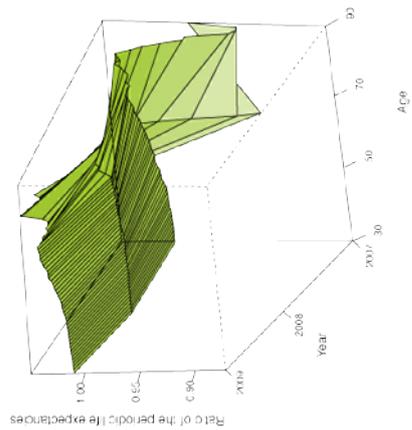
(h) Method 4, female population.

Figure 12: Comparison of the observed and predicted periodic life expectancies.
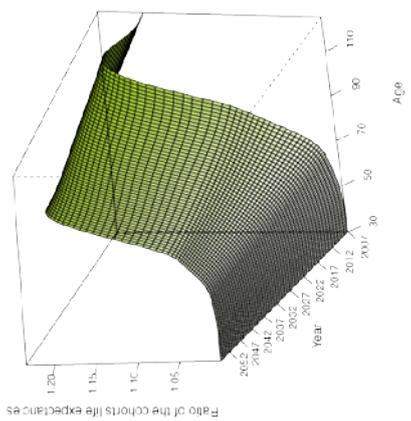
(a) Method 1.
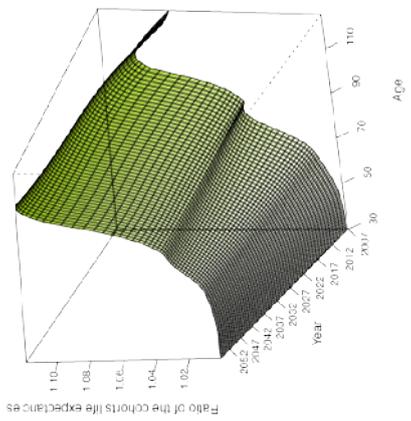
(b) Method 2.
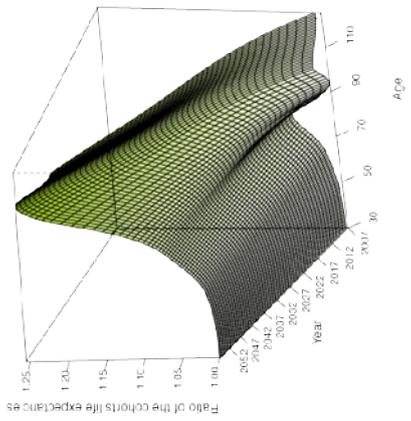
(c) Method 3.
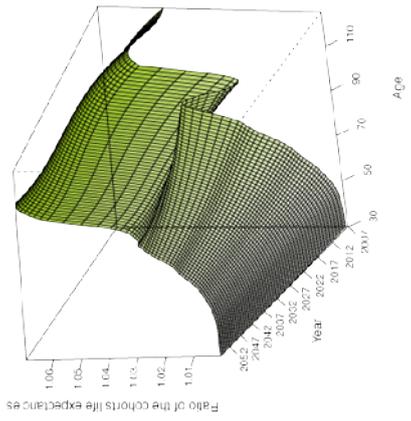
(d) Method 4.
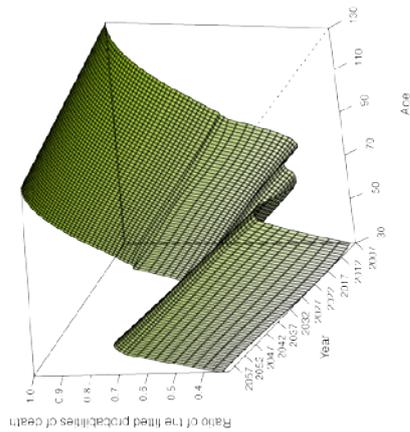
(e) Method 1.
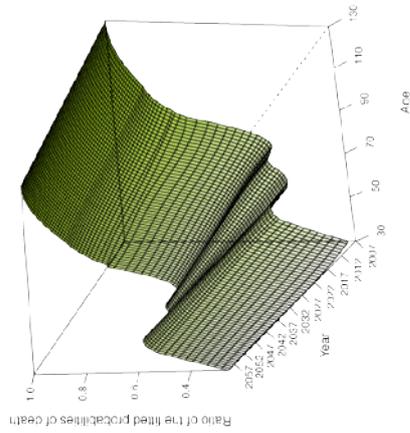
(f) Method 2.
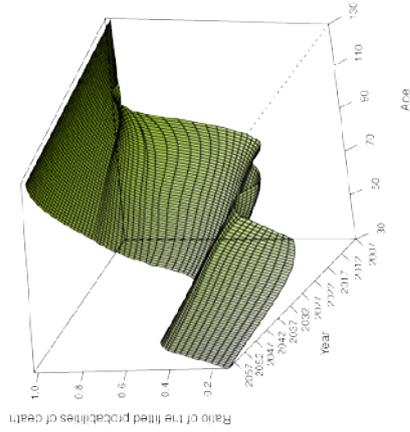
(g) Method 3.
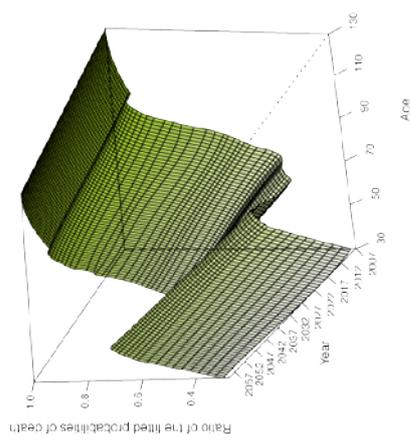
(h) Method 4.

Figure 13: Comparison of the male and female mortality.

Further developments must be considered. We need to assess the overall model risk associated with a prospective mortality table. It would be appropriate to allow for parameters uncertainty to provide a more complete picture of the level of risk on the valuations of an insurer, such as provisioning and capital requirement. Moreover, we have only considered in this study a statical dimension of the adjustment, in relying solely on information available at the construction of the table. We need to take into account a dynamic dimension, allowing a multi-year readjustment of the probabilities of death, based on additional information gathered by the insurer on the current year, such as a sequential change-point detection mechanism providing an indicator about the need for a readjustment.

Acknowledgement

References

Cairns, A. J. G., Blake, D., and Dowd, K. (2006). Pricing death: Frameworks for the valuation and securization of mortality risk. *ASTIN Bulletin*, **36**, 79–120.

Debón, A., Montes, F., and Sala, R. (2006). A comparison of nonparametric methods in the graduation of mortality: Application to data from the Valencia region (Spain). *International statistical Review*, **74**(2), 215–233.

Delwarde, A. and Denuit, M. (2005). *Construction de tables de mortalité périodiques et prospectives*. Assurance Audit Actuariat. Economica.

Delwarde, A., Kachkhdze, D., Olie, L., and Denuit, M. (2004). Modèles linéaires et additifs géneralisés, maximum de vraisemblance local et méthodes relationelles en assurance sur la vie. *Bulletin Français d'Actuariat*, **6**(12), 77–102.

Denuit, M. and Goderniaux, A. C. (2005). Closing and projecting life tables using log-linear models. *Bulletin of the Swiss Association of Bulletin of the Swiss Association of Actuaries*, (1), 29–48.

Felipe, A., Guillén, M., and Pérez-Marín, A. (2002). Recent mortality trends in the Spanish population. *British Actuarial Journal*, **8**(4), 757–786.

Forfar, D., McCutcheon, J., and Wilkie, A. (1988). On graduation by mathematical formula. *Journal of the Institute of Actuaries*, **115**(1), 1–459.

Hunt, A. and Blake, D. (2013). A general procedure for constructing mortality models. *Discussion paper Pension Institute*, (1301), 1–51.

Liddell, F. D. K. (1984). Simple exact analysis of the standardised mortality ratio. *Journal of Epidemiology and Community Health*, **38**, 85–88.

Planchet, F. and Thérond, P. (2011). *Modélisation statistique des phénomènes de durée - Applications actuarielles*. Assurance Audit Actuariat. Economica Paris.

R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Tomas, J. (2011). A local likelihood approach to univariate graduation of mortality. *Bulletin Français d'Actuariat*, **11**(22), 105–153.

Tomas, J. (2013). *Quantifying biometric life insurance risks with non-parametric methods*. Ph.D. thesis, Amsterdam School of Economics Research Institute.

Tomas, J. and Planchet, F. (2013a). Multidimensional smoothing by adaptive local kernel-weighted log-likelihood with application to long-term care insurance. *Insurance: Mathematics & Economics*, **52**(3), 573–589.

Tomas, J. and Planchet, F. (2013b). Construction et validation des références de mortalité de place. Technical report II1291-11 v1.4, Institut des Actuaires.

Tomas, J. and Planchet, F. (2013c). Critères de validation: aspects méthodologiques. Technical report II1291-14 v1.4, Institut des Actuaires.

Tomas, J. and Planchet, F. (2013d). Méthodes de positionnement: aspects méthodologiques. Technical report II1291-12v1.7, Institut des Actuaires.

Tomas, J. and Planchet, F. (2014). Prospective mortality table and portfolio experience. In A. Charpentier, editor, *Computational Actuarial Science, with R*, The R Series, chapter 9. Chapman and Hall.

A   The data

Table 12 presents the observed characteristics of the male and female population.

	Mean age in	Mean age out	Average exposition	Mean age at death	Period of observation Beginning	End
Male pop.	46.54	50.34	3.80	62.43	01/01/2005	31/12/2009
Female pop.	47.53	51.32	3.80	65.24	01/01/2005	31/12/2009

Table 12: Observed characteristics of the male and female population of the portfolio.

Figures 14 et 15 display the observed statistics of the male and female population respectively.

(a) $D_{x,t}$ (b) $E_{x,t}$ (c) $\log q_x(t)$
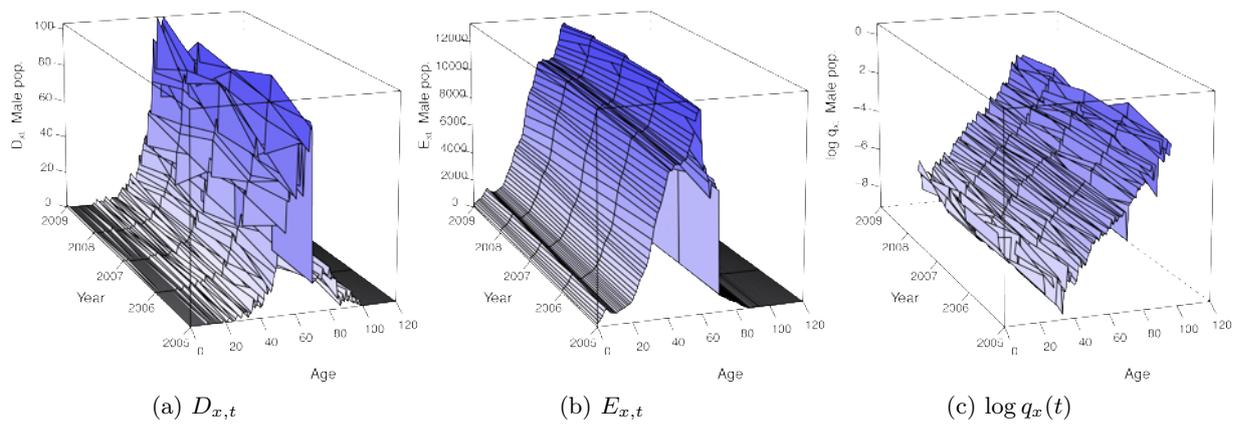
Figure 14: Observed satistics, male population.

(a) $D_{x,t}$ (b) $E_{x,t}$ (c) $\log q_x(t)$
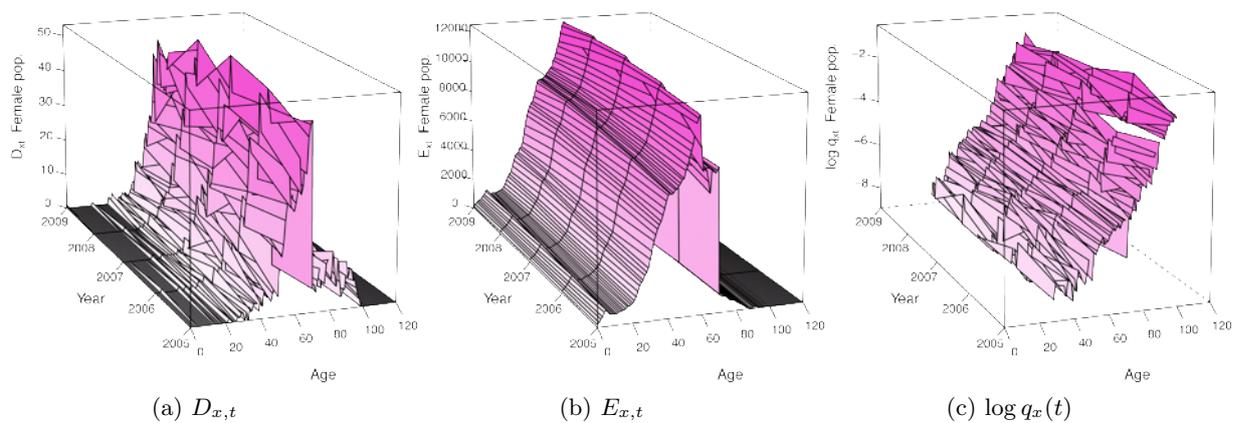
Figure 15: Observed satistics, female population.