



**HAL**  
open science

## Age-Specific Adjustment of Graduated Mortality

Yahia Salhi, Pierre-Emmanuel Thérond

► **To cite this version:**

Yahia Salhi, Pierre-Emmanuel Thérond. Age-Specific Adjustment of Graduated Mortality. *ASTIN Bulletin*, 2018, 48 (2), pp.543-569. hal-01391285

**HAL Id: hal-01391285**

**<https://hal.science/hal-01391285>**

Submitted on 3 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Age-Specific Adjustment of Graduated Mortality\*

Yahia SALHI<sup>(1)</sup> and Pierre-E. THÉRON<sup>(1,2)</sup>

<sup>(1)</sup> Univ Lyon, UCBL, ISFA LSAF EA2429, F-69007, Lyon, France

<sup>(2)</sup>Galea & Associés, 12 avenue du Maine, F-75015 Paris, France

October 20, 2016

## Abstract

Recently, there has been an increasing interest of life insurers to assess their portfolios own mortality risk. The new European prudential regulation, namely Solvency II, emphasized the need to use mortality and life tables that best capture and reflect the experienced mortality, and thus policyholders' proper risk profile, in order to adequately quantify the underlying risk. Therefore, building a mortality table based on the experience from the portfolio is highly recommended and, for this purpose, various approaches have been introduced in the literature. Although, such approaches succeed in capturing the main feature, it remains difficult to assess the mortality when the underlying portfolio lacks of sufficient exposure.

In this paper, we propose to graduate the mortality curve using an adaptive procedure based on the local likelihood, which has the ability to model the mortality patterns even in presence of complex structures and avoid to rely on experts opinion. However, such a technique fails at proposing a consistent yet regular structure when for portfolios with limited deaths. Although the technique borrows the information from the adjacent ages, it is sometimes not sufficient to produce a robust life tables. In presence of such a bias, we propose to adjust the corresponding curve, at the age level, based on a credibility approach. This consists on reviewing, as new observations arrive, the assumption on the mortality curve.

We derive the updating procedure and investigate the benefits of using the latter instead of a sole graduation based on real datasets. Moreover, we look at the divergences in the mortality forecasts generated by the classical credibility approaches including Hardy-Panjer, the Poisson-Gamma model and Makeham framework on portfolios originating from various French insurance companies.

**Keywords:** Credibility; Mortality; Life Insurance; Graduation; Smoothing; Local Likelihood; Prediction

---

\*This work has been supported by the BNP Paribas Cardif Chair "Data Analytics & Models in Insurance" and the ANR project "LoLitA" (ANR-13-BS01-0011). The views expressed in this document are the authors owns and do not necessarily reflect those endorsed by BNP Paribas Cardif.

# 1 Introduction

For insurers, the assessment of the experienced mortality is of paramount importance. The new regulation and norms, Solvency II, shed light on the need of life tables that best reflect the experience of insured portfolios own mortality so as to reliably quantify the underlying risk. Insurers, in France for example, are used to rely on regulatory life tables for pricing and reserving purposes, which are sometimes too conservative. The use of inadequate life tables, being too conservative would considerably affect the financial profitability of the life insurance businesses as well as the insurer competitiveness. Among others, from a Solvency II perspective, using too conservative tables leads to two major impacts: (i) an increase of *Best Estimate* technical provisions (and thus a decrease of *Basic Own-Funds*); (ii) an increase of the *base figure* used for calculating the capital charge for mortality risk.

A natural and straightforward approach to handle this issue is to use the available data at the portfolio level and build an entity-specific mortality table. However, practitioners may face technical difficulties related to the size of the portfolios and the heterogeneity of the guarantees (for the same underlying risk). For instance, an insurer may have a fairly big portfolio but with policyholders holding different insurance contracts: pure endowment contracts, unit-linked contracts with minimum death guarantees, loan insurance and so on. In such a case, it is difficult to build a mortality table based on the sole experience of each product or guarantee. More precisely, the constructed table may not be able to represent the mortality profile of the policyholders thus failing in capturing the underlying risk profile. This should also be the case even if the mortality table is periodically updated with the incoming new data. Especially, the latter may induce significant impacts on the technical reserves if the table has to be updated more frequently over time. Also, if one draws a mortality table only based on the experience stemming from one product or guarantee, she shall have the face a problem of heterogeneity. The latter arises not only at the portfolio level but also for individual ages. In fact, the mortality profile is highly dependent on the age of the individuals and some age groups being *poorly* represented may alter the quantification of the mortality risk at each individual age.

In this paper, we consider an insurer with exposures to different policies and aiming at establishing an experience-based mortality table for each policy and age level, as individuals may differ substantially in their endowment for biometric risk (as showed by some empirical mortality studies, e.g. see [Vaupel et al. \(1979\)](#) and [Hougaard \(1984\)](#) among others). As a first step, we consider a graduation principle to build mortality rates at the insured portfolio level. There are usually two sorts of methods: non-parametric and parametric, see [Forfar et al. \(1988a\)](#) and [Debón et al. \(2006\)](#) for a comprehensive introduction to the use of both graduation techniques. The nonparametric framework is very useful in practice especially when there is sufficient data. This method relies on the use of kernel estimation techniques which were first used for graduation by [Copas and Haberman \(1983\)](#) and [Ramlau-Hansen \(1983\)](#). In the current work, we use local kernel-weighted log-likelihood techniques to graduate the observed mortality with attained age. It was considered in [Tomas \(2011, 2012\)](#), which introduced adaptive smoothing parameters choice with an application in life 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 graduated mortality can be then used as such to project future insurance liabilities related to the underlying population. However, the evolution of the flow of data related to latest available information is not taken into account. This should be, for example, used to update the graduated mortality. However, if one decides to redo a graduation procedure including the new data, the

forecasts are likely to be unstable; adding potential volatility to the underlying reserves and capital charges.

Therefore, the primary contribution of this paper is the incorporation of heterogeneity into the graduated mortality table model by introducing an unobserved variate for individual differences in each attained age. Such an approach has been considered in [Salhi et al. \(2016\)](#) but with different graduated curve. The latter uses a parametric model, i.e. Makeham law, to first built the mortality curve and then applies a credibility to the portfolio-sensitive parameter. Other approach have been also introduced in the literature but work directly on the death counts, e.g. [Hardy and Panjer \(1998\)](#) and [Bühlmann and Gisler \(2005\)](#) among others. Unlike the classical approaches that focus on the update of the aggregate deaths recorded over the whole portfolio, the proposed adjustment approach is intended to enhance the predictive ability of the graduated mortality using a credibility-based revision at the age-level and not on the aggregate portfolio level. By doing so, we implicitly account for the heterogeneity observed at the age-level and has beneficial outputs in pricing and reserving fields when considering the insurance guarantees according to the age.

This note is organized as follows. [Section 2](#) has still an introductory purpose. It specifies the notation and assumptions used in the following. We also describe the graduation procedure based on the local-likelihood approach and recall some statistical inference results used in the sequel. [Section 3](#) introduces the credibility approach to the graduated mortality mortality. We specify the model and make connection with the recent literature and derive the main tools needed to fully characterize the next period prediction of mortality rates when a (multiplicative) credibility factor is taken into account. Finally, [Section 4](#) presents an application with experience data originating from French insurance companies. Finally, some remarks in [Section 5](#) conclude the paper.

## 2 Notation, Assumptions and Preliminaries

**2.1 Notation and Assumptions.** We suppose that we have at our disposal age-specific mortality statistics originating from  $K$  portfolios (or companies) over the period  $[1, T_i]$ , with  $i \in \{1, \dots, K\}$ . For each portfolio  $i$ , we let  $L_{x,t}^i$  denote the number of individuals at attained age  $x$  during calendar year  $t$  and  $D_{x,t}^i$  represents the number of deaths recorded from an exposure-to-risk  $E_{x,t}^i$  measuring the time during which individuals are exposed to the risk of dying. To each of the observations  $j$ , we associate the dummy variable  $\delta_j$  indicating if the individual  $j$  dies or not over one year. That is, for each  $j = 1, \dots, L_{x,t}^i$ , and age-band  $[x, x + 1]$ ,  $\delta_j = 1$  if individual  $j$  dies and null otherwise. Next, we define the time lived by individual  $j$  before her  $(x + 1)^{\text{th}}$  birthday by  $\tau_j$  and assume that we have at our disposal i.i.d. observations  $(\delta_j, \tau_j)$  for each of the  $L_{t,x}^i$  individuals. Then, it is straightforward that:

$$\sum_{j=1}^{L_{x,t}^i} \tau_j = E_{x,t}^i \quad \text{and} \quad \sum_{j=1}^{L_{x,t}^i} \delta_j = D_{x,t}^i, \quad \text{for } i = 1, \dots, K.$$

Furthermore, we assume that the age-specific forces of mortality are constant within bands of time and age, but allowed to vary from one band to the next,  $\varphi_{x+\tau,t+\xi}^i = \varphi_{x,t}^i$  for  $0 \leq \tau < 1$  and  $0 \leq \xi < 1$ . We also denote by  $p_{x,t}^i$  the probability that an individual aged  $x$  in calendar year  $t$  originating from company  $i$  reaches age  $x + 1$ , and by  $q_{x,t}^i = 1 - p_{x,t}^i$  the corresponding probability of death.

Under the assumption of piecewise constant forces of mortality, we have for integer ages  $x$  and calendar years  $t$ ,  $p_{x,t}^i = \exp(-\varphi_{x,t}^i)$  and  $\varphi_{x,t}^i = -\log(p_{x,t}^i)$ . Moreover, the likelihood can be expressed in the following form:

$$\mathcal{L}(\varphi_{x,t}^i) = \exp(-E_{x,t}^i \varphi_{x,t}^i) (\varphi_{x,t}^i)^{D_{x,t}^i}.$$

The associated log-likelihood is then given by

$$\ell(\varphi_{x,t}^i) = \log \mathcal{L}(\varphi_{x,t}^i) = -E_{x,t}^i \varphi_{x,t}^i + D_{x,t}^i \log \varphi_{x,t}^i.$$

Maximizing the log-likelihood  $\ell(\varphi_{x,t}^i)$  gives  $\widehat{\varphi}_{x,t}^i = D_{x,t}^i / E_{x,t}^i$  which coincides with the central death rates  $\widehat{m}_{x,t}^i$ . Then it is apparent that the likelihood  $\ell(\varphi_{x,t}^i)$  is proportional to the Poisson likelihood based on the following assumption

$$D_{x,t}^i \sim \mathcal{P}(E_{x,t}^i \varphi_{x,t}^i). \quad (\text{H0})$$

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 \(2005\)](#). Accordingly, under the assumption of constant forces of mortality between integer values of  $x$  and  $t$ , we consider, henceforth, [\(H0\)](#) to take advantage of the Generalized Linear Models (GLMs) framework.

In the sequel we frequently use the following notation. Let  $D_{x\bullet}^i = \sum_{t=1}^{T_i} D_{x,t}^i$  and  $E_{x\bullet}^i = \sum_{t=1}^{T_i} E_{x,t}^i$  be respectively the company specific time aggregated death counts and exposures. For convenience, we write  $\mathbf{D}^i = (D_{x_1\bullet}^i, \dots, D_{x_n\bullet}^i)^\top$  and  $\mathbf{E}^i = (E_{x_1\bullet}^i, \dots, E_{x_n\bullet}^i)^\top$  the vectors of aggregated deaths and exposures for company  $i$  and also denote  $\boldsymbol{\varphi}^i = (\varphi_{x_1\bullet}^i, \dots, \varphi_{x_n\bullet}^i)^\top$ . Thus, the *bullet* indicates the summation over the corresponding index in the notation introduced above.

**2.2 Local Likelihood Smoothing of Mortality Tables.** The crude estimates of mortality for each company  $i$  might be subject to random fluctuations. Typically, the latter may arise when the underlying exposure to risk is small. Therefore, the common practice in life insurance is whether to graduate mortality using a parametric approach or use some non-parametric smoothing techniques, see [Forfar et al. \(1988a\)](#) and [Debón et al. \(2006\)](#) for a comprehensive introduction to the use of graduation techniques. Here, we consider a non-parametric graduation method of experienced data originating from each company following the approach introduced in [Tomas and Planchet \(2013a,b\)](#). We use the hypothesis of Poisson distributed deaths as in [\(H0\)](#) and assume that the intensity  $\varphi_{x,t}^i$  of the portfolio  $i$  is related to a reference  $\varphi_{x,t}^{\text{ref}}$  as follows:

$$\varphi_{x,t}^i = \exp(f(x)) \varphi_{x,t}^{\text{ref}}, \quad (1)$$

where  $f$  is an unspecified, smooth and deterministic function of the age  $x$ . The latter allows to link the mortality of the company  $i$  to the baseline at the attained age level  $x$ . We define an estimator  $\widehat{\varphi}_{x,t}^i$  of  $\varphi_{x,t}^i$  based on the data from the  $i$ 'th subgroup. To do so, a non-parametric estimation of  $f$  is needed as developed in [Tomas and Planchet \(2013a,b\)](#). More formally, we consider the local kernel-weighted log-likelihood method to estimate the smooth function  $f(x)$ . Such an approach was used to graduate life tables with attained age context in [Delwarde et al. \(2004\)](#), [Debón et al. \(2006\)](#) and [Tomas \(2011\)](#). One of the key advantages of such an approach is its ability to model the mortality patterns even in the presence of complex structures and avoid to rely on experts opinion in graduating mortality, thanks to the flexible functional form of the regression curve. The starting point of our estimation is the specification of the GLM under the assumption [\(H0\)](#) that allows to write the first moment of  $D_{x,t}^i$  as follows:

$$\log \mathbb{E}[D_{x,t}^i] = \log E_{x,t}^i + \log \varphi_{x,t}^i = \log E_{x,t}^i + \log \varphi_{x,t}^{\text{ref}} + f(x), \quad (2)$$

where the term  $\log E_{x,t}^i$  is an *offset*. Such a model has been widely considered in the literature. For example, in [Currie \(2013\)](#), the function  $f$  has the parametric form  $f(x) = \beta_0 + \beta_1 x$  for some

unknown parameters  $\beta_0$  and  $\beta_1$ . In our case, there is no model assumption about  $f$ . The natural extension of this idea is to suppose that for each age  $x_k$ , with  $k = 1, \dots, n$ ,  $f(x_k)$  is a  $p^{\text{th}}$  degree polynomial in  $x_j$ 's, where  $x_j$  is an element in the neighborhood of  $x_k$ . Formally, denoting  $\mathbf{x}_j = (1, x_k - x_j, \dots, (x_k - x_j)^p)^\top$  and  $\boldsymbol{\beta}^i = (\beta_0, \dots, \beta_p)^\top$  we can write  $f(x_k)$  in the following form  $f(x_k) = \mathbf{x}_j^\top \boldsymbol{\beta}^i$ . Next, to estimate  $\boldsymbol{\beta}^i$ , we maximize the likelihood of the model. To do this, we add a weight on the observations at each age. Following Tomas (2012), we use a non-negative weight function, i.e.  $w_j = w_j(x_k)$ , that depends on the distance between the observations and the fitting point  $x_j$ , i.e.

$$\omega_j = \begin{cases} \omega_j(|x_j - x_k|/h), & \text{if } |x_j - x_k|/h \leq 1, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where  $h$  is a smoothing parameter determining the radius of the neighborhood of  $x_k$  used in the smoothing. It gives the bandwidth of the neighbor used in the kernel. For instance, the smallest  $h$  the thinner is the *voisinage* that contributes to the likelihood at each attained age. Examples of the weighing function  $\omega_j$  are given in Tomas (2011, Table 1). In our case we consider a Gaussian kernel and use these weights on the characterization of the likelihood. Accordingly, it follows that the log-likelihood, for portfolio  $i$ , under the assumption (H0) writes:

$$\begin{aligned} L(\boldsymbol{\beta}^i | \mathbf{D}^i, \mathbf{E}^i, (\omega_j)_j) &= \sum_{t=1}^{T_i} \sum_{j=1}^n \omega_j (D_{x_j,t}^i \mathbf{x}_j^\top \boldsymbol{\beta}^i - E_{x_j,t}^i \varphi_{x_j}^{\text{ref}} e^{\mathbf{x}_j^\top \boldsymbol{\beta}^i} - \log D_{x_j,t}^i! + D_{x_j,t}^i \log E_{x_j,t}^i \varphi_{x_j}^{\text{ref}}), \\ &= \sum_{j=1}^n \omega_j (D_{x_j,\bullet}^i \mathbf{x}_j^\top \boldsymbol{\beta}^i - E_{x_j,\bullet}^i \varphi_{x_j}^{\text{ref}} e^{\mathbf{x}_j^\top \boldsymbol{\beta}^i}) + C^i, \end{aligned}$$

where  $C^i$  is a constant offset which does not depend on the parameters  $\boldsymbol{\beta}^i$ . Hence, taking the derivative with respect to each  $\beta^i(p)$ , for  $p = 1, \dots, l$ , leads to the following system of equations,

$$\sum_{j=1}^n \omega_j (x_j - x_i)^p (D_{x_j,\bullet}^i - E_{x_j,\bullet}^i \varphi_{x_j}^{\text{ref}} e^{\mathbf{x}_j^\top \boldsymbol{\beta}^i}) = 0,$$

which can be rewritten in a matrix notation as follows:

$$(\mathbf{X}^i)^\top \mathbf{W}^i (\mathbf{D}^i - \boldsymbol{\varphi}^i) = 0, \quad (4)$$

with  $\boldsymbol{\varphi}^i$  being non-linearly dependent on  $\boldsymbol{\beta}^i$ ,  $\mathbf{X}^i$  is the matrix

$$\mathbf{X}^i = \begin{pmatrix} 1 & x_1 - x_i & (x_1 - x_i)^2 & \cdots & (x_1 - x_i)^n \\ 1 & x_2 - x_i & (x_2 - x_i)^2 & \cdots & (x_2 - x_i)^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x_i & (x_n - x_i)^2 & \cdots & (x_n - x_i)^n \end{pmatrix}, \quad (5)$$

and  $\mathbf{W}^i$  is the weight matrix with elements  $w_{k,j}$  given as  $w_{k,j} = \omega_j$  for  $k = 1, \dots, n$ . The solution of the above equation, i.e. estimations, must be obtained numerically using an iterative algorithm as Nelder-Wedderburn, Newton-Raphson algorithms or the fisher scoring methodology, see Tomas (2011) for further development. From these, we can get the estimation of  $\boldsymbol{\beta}^i$  and  $\boldsymbol{\varphi}^i$  denoted, henceforth, by  $\widehat{\boldsymbol{\beta}}^i$  and  $\widehat{\boldsymbol{\varphi}}^i$

**2.3 Inference of the Graduated Mortality.** The aim of this subsection is to characterize the statistical feature of the estimators considered above. We recall some results from [Tomas \(2011\)](#) well known in the nonparametric smoothing literature, see e.g. [Tibshirani and Hastie \(1987\)](#) and [Wand and Jones \(1994\)](#), regarding particularly the variance of the graduated mortality and the expected behavior of these estimations. Accordingly, using theoretical results concerning bias and variance the estimator  $\hat{\varphi}^i$  is shown to be asymptotically design-unbiased and consistent. It is, for instance, recalled in [Tomas \(2011\)](#) that the smoothed mortality rates  $\hat{\varphi}^i$  are unbiased estimators of  $\varphi^i$  in the sense that:

$$\mathbb{E}[\hat{\varphi}^i] \approx \varphi^i. \quad (6)$$

The above approximation is based on the inspection of the mean squared errors, which are commonly used to assess the bias of the estimation in such a framework. Expressions of the latter are available in the classical textbooks and the readers is referred to ([Tomas 2011](#), Sec. 3.2.) which provides an approximation to the bias of the estimator  $\hat{\varphi}^i$ . Unlike the linear model fitting, there is no exact expression for moments of  $\hat{\varphi}_x^i$  due to the non-linearity in [Equation \(4\)](#). Using a multivariate version of Taylor series expansion around  $\beta^i$  allows to use classical results on the inference of the estimated parameter  $\hat{\beta}^i$ . Note that this approximation depends on the bandwidth of the neighborhood  $h$  used in the kernel. More precisely, the bias is decreasing with the bandwidth. This is, particularly, reasonable in practice, because a large bandwidth induces a miss-fitting of the local polynomials and hence also the sum of squared residuals. To derive the second order moment of  $\hat{\varphi}^i$ , a variance approximation based on Taylor linearization is also generally suggested and shown to be consistent, see [Loader \(2006\)](#). More precisely, we have the following expression for the variance:

$$\text{Var}(\hat{\varphi}^i) = \hat{\varphi}^i \otimes \mathbf{S}^i \mathbf{S}^{i\top}, \quad (7)$$

with " $\otimes$ " being the pointwise division operator. Here, the matrix  $\mathbf{S}^i$  is given as follows:

$$\mathbf{S}^i = \begin{pmatrix} s_1^i(x_1) & s_2^i(x_1) & s_3^i(x_1) & \cdots & s_n^i(x_1) \\ s_1^i(x_2) & s_2^i(x_2) & s_3^i(x_2) & \cdots & s_n^i(x_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_1^i(x_n) & s_2^i(x_n) & s_3^i(x_n) & \cdots & s_n^i(x_n) \end{pmatrix}, \quad (8)$$

with rows  $\mathbf{s}^i(x_j)^\top = (s_1^i(x_j), s_2^i(x_j), \dots, s_n^i(x_j)) = (\mathbf{X}^{i\top} \mathbf{W}^i \mathbf{X}^i)^{-1} \mathbf{X}^{i\top} \mathbf{W}^i$ , where  $\mathbf{W}^i$  is the weight matrix and  $\mathbf{X}^i$  is given in [Equation \(5\)](#)

**2.4 Small Sized Portfolios and Heterogeneity.** It is worth mentioning that the relational (proportional) model considered in [Equation \(1\)](#) implicitly accounts for differential mortality that may arise due to portfolio specific features, e.g. particular socioeconomic groups involved, income level, etc. Indeed, insured portfolios show a typical behavior compared to a national mortality. The mortality of insured population is significantly lower than the national population from which it is drawn. However, when it comes to the study of the mortality at a single portfolio level, some stylized facts arise, which might compromise the efficiency of the graduation procedure. For instance, insured population are generally of small size, so none or very few deaths are observable at some ages. Therefore, the use of the model [\(1\)](#) and the local-likelihood based estimation procedure advocates using the information stemming from the adjacent ages to construct the mortality curve. This *learning* procedure will enhance the determination of the mortality at a given age. However, when successive ages lack of information, the approach exposed above will need a large bandwidth  $h$  for the

estimator to access distant ages with sufficient and reliable information. By doing so, we increase the bias surrounding the smoothed curve. In fact, as noted above, the mean of squared errors measuring the bias due to the local regression increase with the bandwidth  $h$ .

Due to these different sources of uncertainty we suppose that the *true* mortality curve  $\varphi_x^i$ , for  $x = x_1 \cdots, x_n$ , is known up to an unobservable multiplicative factor  $\Theta_x^i$ . In other words, the portfolios examined should be regarded as a sample of the reference. Estimates based on the data will be subject to sampling errors and the smaller the group is, the greater will be the relative random errors in the number of deaths and the less reliable will be the resulting estimates. This argument is extended to include the bias stemming at the attained age level due the consideration exposed above. Thus, if one has estimated the curve using the non-parametric approach, the *true* curve is an adjustment of the latter as multiplied by the random and non-observable parameter  $\Theta_x^i$ . This reasoning is closely related to the multiplicative frailty model that was introduced by [Vaupel et al. \(1979\)](#). This concept is well established in biostatistics and survival analysis, see e.g. [Hougaard \(1984\)](#) and [Aalen \(1988\)](#) among others. In the actuarial literature frailty models have also been used. An overview of the use of frailty models to handle the heterogeneity phenomenon in life insurance is given in [Olivieri \(2006\)](#). In our case, we address this problem using the credibility theory, see [Bühlmann and Gisler \(2005\)](#) for more details.

### 3 Company-Specific Relative Risk Level

We suppose that there are  $K$  companies, with a portfolio of observed individuals of ages ranging from  $x_1$  to  $x_{n_i}$ . Here, the  $n_i$ 's might be all different to be in line with the insurance practices. This kind of information structure is similar to the so-called *unbalanced* framework used in actuarial science. For the sake of readability, without loss of generality, we will henceforth assume similar observed age groups for all companies, i.e.  $n = n_1 = \cdots = n_K$ . With a slight variation to the model, however, it can be easily extended to the unbalanced case.

**3.1 The credibility model.** Given the specific parameterization of the problem, one may think of the  $K$  portfolios as a subset of the reference population and thus each population is characterized by a risk profile. The latter is due to the heterogeneous sizes of the portfolios as well as the underlying guarantees (for the same underlying risk). These sources of heterogeneity might also induce an age varying risk profile within the same portfolio. Therefore, for a company  $i$ , we let the vector  $\Theta^i = \text{diag}(\Theta_{x_1}^i, \cdots, \Theta_{x_n}^i)$  be her relative risk level. For  $x \in \{x_1, \cdots, x_n\}$ , each  $\Theta_x^i$  characterizes the age-specific risk level, which are unobservable random variables.

The primal objective is to characterize the force of mortality of each company  $i$  at a specific age  $x$  through the proportional relationship introduced in [Subsection 2.2](#), i.e.

$$\varphi^i = \Theta^i \alpha, \tag{1}$$

where  $\alpha = (\alpha_{x_1}, \cdots, \alpha_{x_n})^\top$  such that for  $j = 1, \cdots, n$ , we have  $\alpha_{x_j} = f(x_j)\varphi_{x_j}^{\text{ref}}$ . This model suggests that for each company  $i$ , the age-specific experienced force of mortality varies around the baseline  $\alpha_x$ , which can be seen as a reference or best-estimate mortality. This fluctuation is modeled by a heterogeneity parameter  $\Theta_x^i$  capturing the individual properties (heterogeneity) of each company at attained age  $x$ . Thus, using new incoming data should allow to update the the next period mortality  $\varphi_x^i$  by adjustment following model in Equation (1). The approach is to first find an estimator  $\hat{\varphi}_x^i$  of  $\varphi_x^i = \Theta_x^i \alpha_x$  for each company  $i$  using the likelihood-based approach introduced in [Section 2](#).



Henceforth, the notation  $\widehat{\varphi}_x^i | \Theta_x^i$  refers to the estimation of the quantity  $\alpha_x = f(x)\varphi_x^{\text{ref}}$ , which, by abuse of language, referred to as the estimated mortality conditional on the risk profile. [Nielsen and Sandqvist \(2000\)](#) and [Gustafsson et al. \(2006\)](#) proposed a similar framework to quantify the operational risk, where an adequate quantity related to losses from different lines of business share a common baseline. In our setting, we generalize their approach to the case where the mortality is structured by age (line of business) and intensities of mortality are dependent within the same group. Indeed, the random variables  $\widehat{\varphi}_{x_1}^i, \dots, \widehat{\varphi}_{x_n}^i$  are assumed to be dependent, namely, the force of mortality of one age does directly impact those of other ages. This is mainly due to the graduation of mortality at a given age, which weights up over the adjacent age groups, see [Section 2](#). This dependency will be explored later on this section. Finally, in order to characterize the next period mortality level, we make use of the credibility theory. For this purpose and using the usual credibility setting, we shall make the following assumptions:

(A1) The random vectors  $\Theta^i$  are independent across companies and ages. Moreover, for  $i = 1, \dots, K$ ,  $\Theta_x^i$ 's are identically distributed with  $\mathbb{E}[\Theta^i] = \mathbf{I}_n$  and  $\text{Var}(\Theta^i) = \sigma$ , where  $\sigma$  is a diagonal matrix with elements  $\sigma_x$  and  $\mathbf{I}_n$  is the identity matrix.

(A2) The random vectors  $(\varphi^i, \Theta^i)$ ,  $i = 1, \dots, K$ , are independent across companies

(A3)  $\varphi_{x_1}^i, \dots, \varphi_{x_n}^i$  are conditionally independent given  $\Theta^i$ .

The first assumption (A1) ensures that the baseline mortality produces the a priori expected number of deaths under the model assumption (H0), in the sense that  $\mathbb{E}[D_{x,\bullet}^i] = \mathbb{E}[\Theta_x^i \alpha_x] = \alpha_x$ . The assumptions (A2) means that the risk profiles are independent over portfolios. In other words, the successive realizations of the mortality intensity (so as the death counts) for any portfolio are independent of each other except through the risk parameter. Finally, assumption (A3) translates the dependency of the mortality over ages. The latter is only captured by the vector  $\Theta^i$ . Conditionally on the latter the forces of mortality at the age level are independent.

As noted before,  $\widehat{\varphi}^i | \Theta^i$  is the *conditional* local-likelihood estimator of the intensity in Equation (1) based on the data from the  $i^{\text{th}}$  portfolio as developed in [Section 2](#). In view of the the assumptions (A1)-(A3), it is important to recall that conditional on the knowledge of the risk profile  $\Theta^i$  the theoretical properties of  $\widehat{\varphi}^i$  are identical to those the of local-likelihood estimator considered in [Subsection 2.3](#). This will be used, among others, in the following lemma, in order to state some fundamental features of the dependence structure.

[A revoir la notation pour le produit]

**Lemma 1** *Under assumptions (A1), (A2) and (A3) and the notation above, we have*

(i) *The first order moment of  $\varphi^i$  is given by*

$$\mathbb{E}[\varphi^i] = \alpha. \quad (2)$$

(ii) *The variance matrix of  $\widehat{\varphi}^i | \Theta^i$ , denoted  $\Sigma^i(\Theta^i) = \text{Var}(\widehat{\varphi}^i | \Theta^i)$ , is given by*

$$\Sigma^i(\Theta^i) = \varphi^i \mathbf{S}^i \mathbf{S}^{i\top}. \quad (3)$$

*Hence, the variance  $\Sigma^i = \text{Var}(\widehat{\varphi}^i)$  can be written as*

$$\Sigma^i = (\mathbf{S}^i \mathbf{S}^{i\top} + \sigma) \alpha \quad (4)$$

(iii) The covariance of  $\varphi_x^i$  with  $\widehat{\varphi}_x^i$  is given by

$$\text{Cov}\left(\varphi_x^i, \widehat{\varphi}_x^i\right) = (\sigma_x \alpha_x^i)^2 \mathbf{e}_{\delta_x}, \quad (5)$$

with  $\delta_x = j$  if  $x = x_j$  and  $\mathbf{e}_j$  is the vector with all 0's except for a 1 in the  $j^{\text{th}}$  coordinate.

**Proof.** To show these results, we make an intensive use of the law of total variance.

(i) Equation (2) is a direct consequence of assumption (A1) which gives  $\mathbb{E}[\varphi^i | \Theta^i] = \Theta^i \alpha$ .

(ii) The conditional variance  $\Sigma^i(\Theta^i)$  is directly derived from the calculus in Subsection 2.3. Hence, to check (4), the law of total variance gives

$$\begin{aligned} \Sigma^i &= \mathbb{E}[\text{Var}(\widehat{\varphi}^i | \Theta^i)] + \text{Var}(\mathbb{E}[\widehat{\varphi}^i | \Theta^i]), \\ &= \mathbb{E}[\widehat{\varphi}^i] \otimes \mathbf{S}^i \mathbf{S}^{i\top} + \alpha \otimes \text{Var}(\Theta^i) = \alpha \otimes (\mathbf{S}^i \mathbf{S}^{i\top} + \sigma). \end{aligned}$$

(iii) Finally, to prove (5), notice that  $\text{Cov}(\varphi_x^i, \widehat{\varphi}_x^i | \Theta^i) = 0$ . Thus,

$$\begin{aligned} \text{Cov}(\varphi_x^i, \widehat{\varphi}_x^i) &= \text{Cov}(\mathbb{E}[\varphi^i(x) | \Theta^i], \mathbb{E}[\widehat{\varphi}^i | \Theta^i]) + \mathbb{E}[\text{Cov}(\varphi_x^i, \widehat{\varphi}_x^i | \Theta^i)], \\ &= \text{Cov}(\Theta_x^i \alpha_x^i, \Theta^i \otimes \alpha) = (\sigma_x \alpha_x^i)^2 \mathbf{e}_{\delta_x}, \end{aligned}$$

where the last equality follows from the independence assumption in (A1). ■

**3.2 The next-period linear per-age mortality estimator.** The goal is to predict the future force of mortality for each company  $i$  at the age level. Therefore, we will be looking for the inhomogeneous credibility predictor corresponding to the linear estimators of  $\varphi_x^i$ . We thus solve the following optimization problems:

$$\min_{c_{0,x}^i, \mathbf{c}_x^i} \mathbb{E} \left[ \left( \varphi_x^i - c_{0,x}^i - \mathbf{c}_x^{i\top} \widehat{\varphi}^i \right)^2 \right], \quad (6)$$

where  $c_{0,x}^i \in \mathbb{R}$  and  $\mathbf{c}_x^i \in \mathbb{R}^n$ . This formulation suggests adjusting the next period force of mortality at a given age using the information stemming from the other age groups. This should enhance the prediction for ages with low or sparse information using the credibility in ages of high information. Based on Proposition 1, we can easily derive the inhomogeneous credibility estimators of  $\varphi^i$ . Indeed, we state the following proposition.

**Proposition 1** *The point estimate of the linear factors in (6) can be written as follows*

$$c_{0,x}^i = (\mathbf{1}_n - \mathbf{c}_x^i)^\top \alpha \quad \text{and} \quad \mathbf{c}_x^i = (\sigma_x \alpha_x^i)^2 (\Sigma^i)^{-1} \mathbf{e}_{\delta_x}. \quad (7)$$

The next period predicted mortality (estimator)  $\widetilde{\varphi}^i$  of  $\varphi^i$  is given by

$$\widetilde{\varphi}^i = (\mathbf{I}_n - (\alpha (\Sigma^i)^{-1} \sigma \alpha)^\top) \alpha + (\alpha (\Sigma^i)^{-1} \sigma \alpha)^\top \widehat{\varphi}^i. \quad (8)$$

**Proof.** Let us first derive the intercept  $c_{0,x}^i$ . To do this, we develop the expectation in Equation (6) and take the derivative with respect to  $c_{0,x}^i$ . This yields to the following equality:

$$c_{i,0} + (\mathbf{c}_x^i)^\top \mathbb{E}[\widehat{\varphi}^i] = 1.$$

On the other hand, differentiating the expectation in Equation (6) with respect to the vector  $\mathbf{c}_x^i$

gives rise to the following variance

$$\text{Var}\left(\varphi_x^i - (\mathbf{c}_x^i)^\top \widehat{\varphi}^i\right),$$

needed to fully characterize the solution. The latter can be computed using results in [Lemma 1](#). Indeed, we can write:

$$\text{Var}\left(\varphi_x^i - (\mathbf{c}_x^i)^\top \widehat{\varphi}^i\right) = \text{Var}(\varphi_x^i) - 2(\mathbf{c}_x^i)^\top \text{Cov}(\varphi_x^i, \widehat{\varphi}^i) + (\mathbf{c}_x^i)^\top \Sigma^i (\mathbf{c}_x^i)^\top.$$

Taking the derivative with respect to the vector  $(\mathbf{c}_x^i)^\top$  yields

$$2\text{Cov}(\varphi_x^i, \widehat{\varphi}^i) - 2\Sigma^i \mathbf{c}_x^i = 0.$$

The terms  $\Sigma^i$  and  $\text{Cov}(\varphi_x^i, \widehat{\varphi}^i)$  are given in [Lemma 1](#), which concludes the proof.  $\blacksquare$

Note that we are able to estimate all the components needed to characterize the next-period intensity  $\widetilde{\varphi}^i$ , except for the variance  $\sigma^i$ . Remarking that  $\widehat{\varphi}^i$  is an estimator of  $\varphi^i = \alpha \otimes \Theta^i$ , we can write  $\widehat{\Theta}^i = \widehat{\varphi}^i \oslash \alpha$ , with "  $\oslash$  " being the pointwise division operator. Therefore, a natural choice of the estimator of  $\sigma$  is

$$\widehat{\sigma} = (\widehat{\Theta}^i - \mathbf{I}_n)^\top (\widehat{\Theta}^i - \mathbf{I}_n). \quad (9)$$

We can now derive the

**Lemma 2** *The optimal credibility estimator of  $\Theta^i$  is given by*

$$\widetilde{\Theta}^i = (\mathbf{I}_n - (\alpha(\Sigma^i)^{-1}\sigma\alpha)^\top) \mathbf{1}_n + (\alpha(\Sigma^i)^{-1}\sigma\alpha)^\top \widehat{\Theta}^i,$$

and the next period prediction of  $\varphi^i$  can be approximated by  $\alpha \otimes \widetilde{\Theta}^i$ .

**Remark 1** *The adjustment procedure described in [Proposition 1](#) and [Lemma 2](#) can be written for each individual age  $x$  in the following form:*

$$\widetilde{\varphi}_x^i = (1 - z_x^i)\alpha_x + z_x^i \widehat{\varphi}_x^i,$$

where  $z_x^i$  is the credibility factor that can be written in the following form:

$$z_x^i = \frac{(\alpha_x)^2 \sigma_i^2}{(\alpha_x)^2 \sigma_i^2 + \widehat{\varphi}_x^i \|\mathbf{s}^i(x)\|^2}.$$

Here, recall that  $\|\mathbf{s}^i(x)\|^2 = \sum_{j=1}^n (s_j^i(x))^2$  and measures the reduction in variance of the smoothed mortality curve  $\widehat{\varphi}_x^i$ .

**Remark 2** *All the ingredients required to implement the credibility approach in [Lemma 2](#), in order to predict the next-period estimator, are already determined. However, we still need to characterize an estimation for  $\alpha$ . To do this, we borrow the same procedure considered in [Nielsen and Sandqvist \(2000\)](#), which amend to estimate  $\alpha$  as linear weighted average over the portfolios.*

## 4 Numerical Analysis

**4.1 Source of Data.** The data come from studies conducted by *Institut des Actuaire*s. These studies include in total 14 portfolios covering the period 2007–2011 with each companies contributing

data for at least 4 of a possible 5 years. [Table 1](#) presents the observed characteristics of the male population of the portfolios. For this dataset, we are considering respectively  $T_i = 3$  and  $T_i = 4$  for all companies. The remaining years serve to test the predictive feature of the model through an in-sample analysis. The considered analysis follows similar lines as in [Salhi et al. \(2016\)](#), which also exploits the same dataset. Therefore, the age band for all companies ranges from  $x = 30$  to 95 years old. [Figure 1](#) shows the age distribution of the portfolios, i.e. the aggregate number of individuals exposed to risk at each attained age. It graphically depicts the heterogeneity observed between the portfolios with insured holding different policies. On the other hand, the baseline mortality  $\varphi_x^{\text{ref}}$  is a market table, denoted IA2013<sup>1</sup>, constructed for the French insurance market provided by *Institute des Actuaires*, see [Tomas and Planchet \(2013c\)](#). Before proceeding to the implementation of the

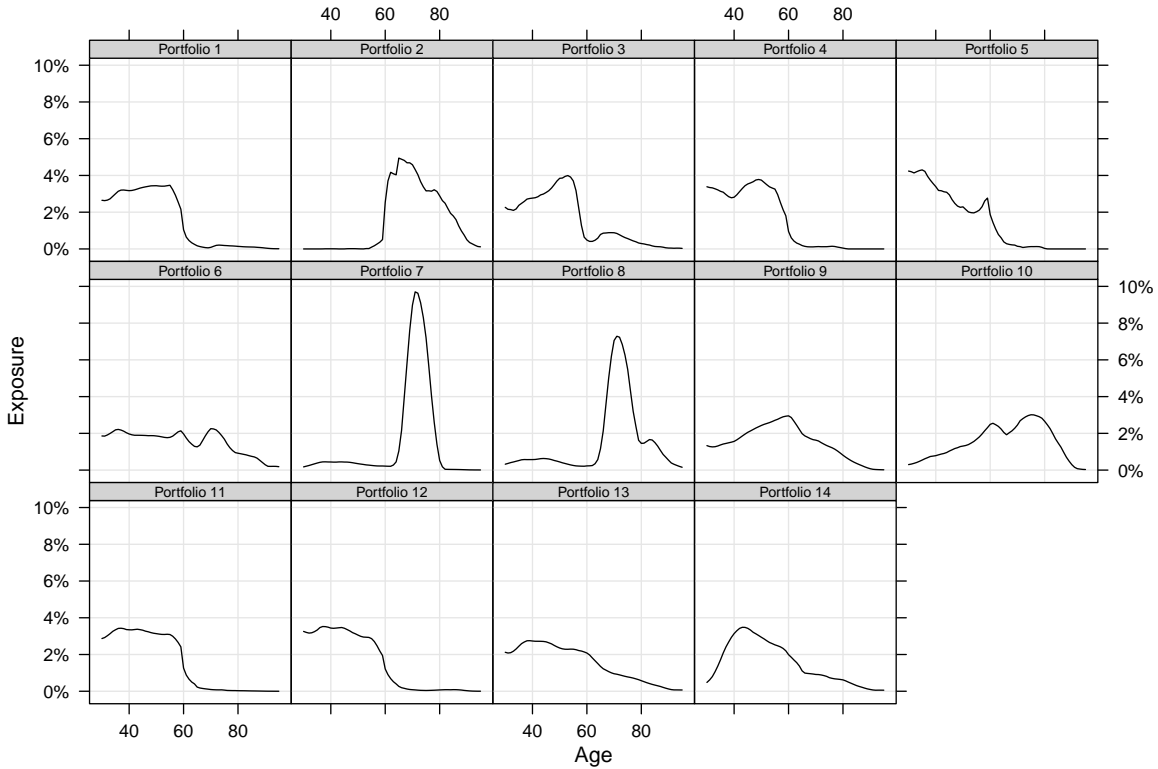
**Table 1:** *Observed characteristics of portfolios population.*

	Period of observation		Mean age		Average exposure	Mean age at death	Size
	Beginning	End	In	Out			
<b>P1</b>	01/01/07	12/31/11	36.96	39.74	2.77	68.78	616390
<b>P2</b>	01/01/07	12/31/11	69.3	73.35	4.05	80.34	7589
<b>P3</b>	01/01/07	12/31/10	40.16	43.1	2.94	71.77	80086
<b>P4</b>	01/01/07	12/31/11	37.5	41.13	3.63	54.08	93165
<b>P5</b>	01/01/07	12/31/11	36.9	39.1	2.2	59.31	21540
<b>P6</b>	01/01/07	12/31/10	48.5	52.11	3.62	82.34	847469
<b>P7</b>	01/01/07	12/31/11	66.65	71.29	4.64	73.68	89507
<b>P8</b>	01/01/07	04/13/11	67.51	71.38	3.86	80.72	78650
<b>P9</b>	01/01/07	06/30/11	45.97	49.6	3.62	73.17	1556150
<b>P10</b>	01/01/07	12/31/11	62.97	67.64	4.67	79.77	132990
<b>P11</b>	01/01/07	12/31/11	38.89	42	3.11	56.44	420405
<b>P12</b>	01/01/07	12/31/11	37.05	39.2	2.15	57.41	904020
<b>P13</b>	01/01/07	12/31/11	43.01	46.89	3.88	71.03	848757
<b>P14</b>	01/01/07	12/31/11	50.12	54.16	4.04	72.37	233488

methodology developed in the previous sections, we must should look deeper into the particular feature of our dataset.. Specifically, we must focus on those that may arise specific concerns when it comes to the graduation of a mortality table using the smoothing procedure considered in [Section 2](#). As previously reported, the experienced mortality does not only suffer from a small sample size but also the under-representation of those within some age groups. This is typically the case of portfolio P2, see [Table 1](#) and [Figure 1](#). In fact, we have a small size of 7589 individuals with only 2% aged under 60. This is the case for portfolios P7, P8 and to some extent P10, but with a bigger exposure. For these portfolios, the use of the smoothing procedure in [Subsection 2.2](#) has the advantage of borrowing the information in age bands where the exposure is substantially larger. However, this may alter the local properties generally required for a mortality curve. Indeed, enlarging the smoothing window  $h$ , giving access to far distant ages, may amend the increasing mortality over ages, which is not only a very much sought behavior but also a biologically reasonable quality.

**4.2 Entity-Specific Graduated Mortality.** In order to implement the likelihood based graduation approach, we need to identify the fitting variables. To this end, we first choose a weight function

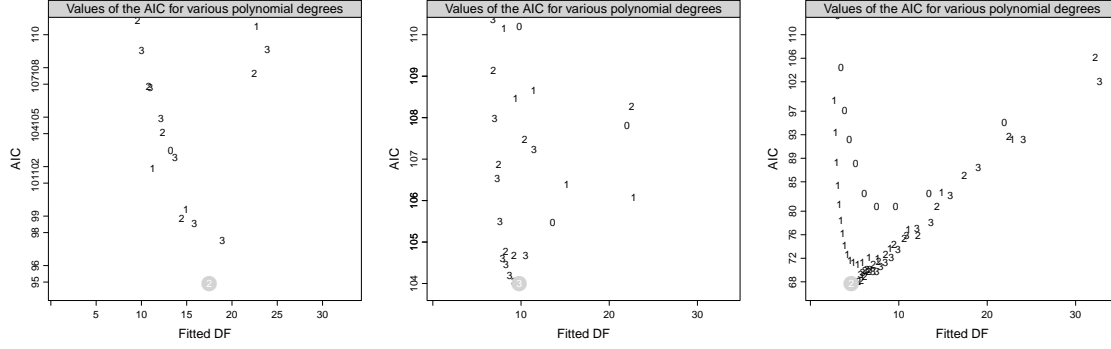
<sup>1</sup>This table is derived from mortality trends originating from the INSEE table, French national bureau of statistics, and covers the period 2007-2060.



**Figure 1:** *Distribution of age groups in the portfolios.*

More precisely, we should characterize the smoothing parameters  $h$  and  $p$  referring, respectively, to (i) the window-width or bandwidth  $h$  which determines how fast the weights decrease and (ii) the degree of the polynomial used for smoothing. To do so, we have to balance the trade-off between bias and variance on the estimation of the mortality as they move on the opposite direction in terms of the smoothing parameters. Formally, we compute a number of candidate fits and use a criteria to select, among the fits, the one with the lowest score. For instance, we use a generalization of the Akaike information criterion (AIC) based on the deviance function taking into account the likelihood as well as the degrees of freedom of the fitting model. The smoothing parameters are selected based on graphical diagnostics. [Figure 2](#) displays the AIC scores against the fitted degrees of freedom for the portfolios 1, 6 and 14. The interpretation of such a graphic is as follows. A two-degree 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. In the left panel of [Figure 2](#), the graphic represents the selection procedure for the portfolio P1. This consists of scores for various polynomial degrees. The optimal choice of the smoothing parameters corresponds to the point on the plan where with the lowest AIC score. For portfolio P1, the lowest score corresponds to a cubic fit with 5 degrees of freedom leading to a smoothing window of 7 points. In [Table 2](#), we report the optimal parameters for each portfolio.

In [Table 2](#), we can see that for some portfolios the optimal choice of the input controls induce high level of degrees of freedom, i.e. portfolios P1 and P5. This is to say that the corresponding "smoothed" curves  $\varphi^i, i = 1, 5$ , will be noisy showing many feature. Indeed, the degree of freedom is a qualitative proxy for the regularity of graduated mortality curve as the *smoothness* inversely to



**Figure 2:** AIC scores for various polynomial degrees for portfolios P1, P6 and P14

**Table 2:** Local-likelihood smoothing parameters' optimal choice

	P1	P2	P3	P4	P5	P6	P7
<b>AIC</b>	98.90	29.34	79.68	78.91	61.10	106.73	68.37
<b>Degree</b>	2	1	1	1	3	3	1
<b>DF</b>	18	3	3	6	16	10	8
	P8	P9	P10	P11	P12	P13	P14
<b>AIC</b>	74.11	83.96	53.44	70.86	74.82	82.70	78.37
<b>Degree</b>	2	1	2	1	1	3	2
<b>DF</b>	4	7	5	6	6	7	5

the degree of freedom. This feature can already be deduced from the limited amount of information (exposures) that are at our disposal for these portfolios. For the remaining portfolio, the degrees of freedom are relatively small. In the following, we will implement the credibility approach described in the last section to assess the impact of the latter on the graduated mortality curves.

**4.3 Next-Period Mortality Rate.** Here, we consider the mortality experience over the period 2007 – 2010 upon which we calibrate the smoothing procedure considered in the above subsection. For each portfolio, we build a graduated mortality table  $\hat{\varphi}^i$  and aim at adjusting the latter for the next period projection. For each age  $x$ , the graduated mortality gives a candidate rate for the next period, i.e.  $\varphi_x^i$ . The insurer has the possibility of whether to rely on this rate or adjust the latter given the experience stemming from the other rates at other ages. In other words, the mortality used for the next period forecasts can be adjusted using the credibility formula in Equation (8). To do this, we estimate the different quantities needed to implement (8) as follows:

- (i) Following Remark 2, the expected mortality rate  $\alpha$  can be estimated as follows:

$$\hat{\alpha} = \left( \sum_{x=x_1}^{x_n} \sum_{i=1}^d E_{x,\bullet}^i \varphi_x^i \right) / \sum_{i=1}^d E_{\bullet,\bullet}^i.$$

- (ii) The weight loading matrix  $S^i$  is given as an output of the graduate step and can be estimated using Equation (8).

- (iii) The diagonal matrix  $\sigma^i$  relies on the variance of  $\Theta_x^i$ 's which might be estimated given that

$\widehat{\Theta}_x^i = \widehat{\varphi}_x^i / \widehat{\alpha}$ , and thus we can write

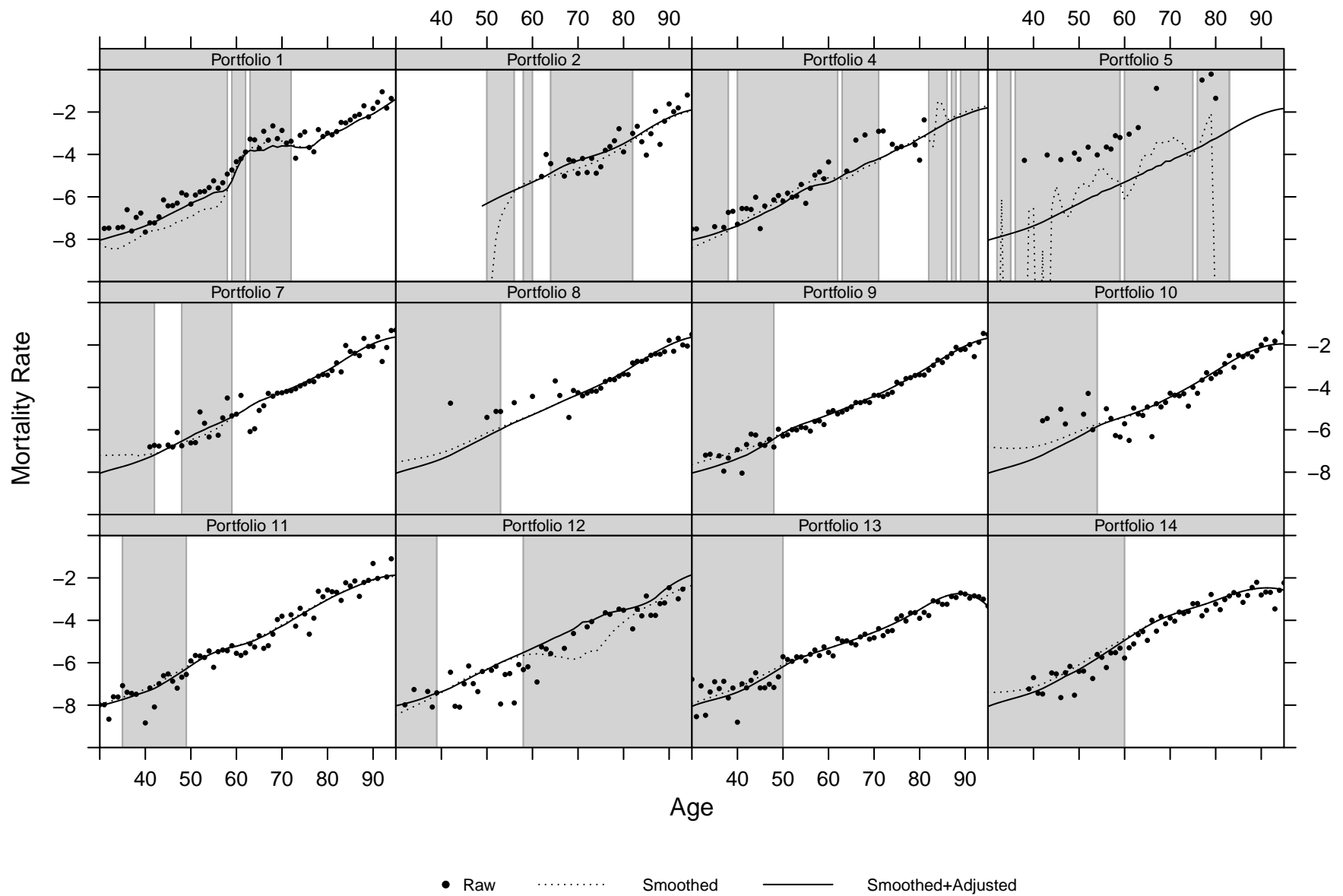
$$\sigma_x^i = \sum_{i=1}^d \left( \widehat{\varphi}_x^i / \widehat{\alpha} \right)^2 / d - \left( \left( \sum_{i=1}^d \widehat{\varphi}_x^i / \widehat{\alpha} \right) / d \right)^2.$$

Figure 3 depicts, respectively, the graduated mortality over the period 2007 – 2010 as described above and the next-period (2011) mortality rates using the credibility formula in Equation (8). Similarly, Figure 4 represents the next-period predicted deaths using the two mortality rates. In these figures we grayed areas (ages) where the relative difference between the smoothed mortality and its adjusted counterpart exceeds a 10% level. More precisely, this corresponds to the ages  $x$  where  $|\widetilde{\varphi}_x^i - \widehat{\varphi}_x^i| / \widehat{\varphi}_x^i > 0.1$ . At a first glance, we remark that the credibility adjustment does change the mortality rate and overall propose a smoother curve compared to the initial one, and this is even evident when dealing with portfolios with small portfolios and high degrees of freedom. In fact, when we deal with portfolios such as P5, where the exposure-to-risk as well as the underlying deaths are very limited, the smoothing approach fails to capture the mortality structure and the output of the procedure proposed in Section 2 are very irregular and noisy. Indeed, as noted above such a procedure need information stemming from adjacent ages when a particular age lacks of sufficient exposure. The case of P5 is very appealing of the limit of the non-parametric smoothing as the limitation on the information is shared over the ages. This is why the corresponding degree of freedom is high and the AIC is low, see Table 2, and explains the irregular curve (dashed line) for the smoothed mortality.

On the other hand, the degrees of freedom given as  $\text{tr}(\mathbf{S}^i \mathbf{S}^{i\top})$  provide information on the credibility of the smoothed curve  $\widehat{\varphi}^i$ . In fact, as we can see in Equation (8) or in a more tractable way as in Remark 1, the higher the degrees of freedom, i.e.  $\text{tr}(\mathbf{S}^i \mathbf{S}^{i\top}) = \sum_{x=x_1}^{x_n} \|\mathbf{s}^i(x)\|^2$ ; the smallest is the weight attached to the smoothed curve  $\varphi^i$  (in aggregate). At the age level, one should look the component-wise of the degrees of freedom, i.e. the individual variance  $\|\mathbf{s}^i(x)\|^2$ . That being said, we can conclude that the parameter driving the adjustment at the age levels is vanishing meaning that the adjusted mortality rate  $\widetilde{\varphi}_x^i$  is close to the reference  $\alpha_x$ . It comes as no surprise, then, to find that the adjusted curve tends to obliterate this undesired effect thanks also to the information coming from other ages but from different portfolios.

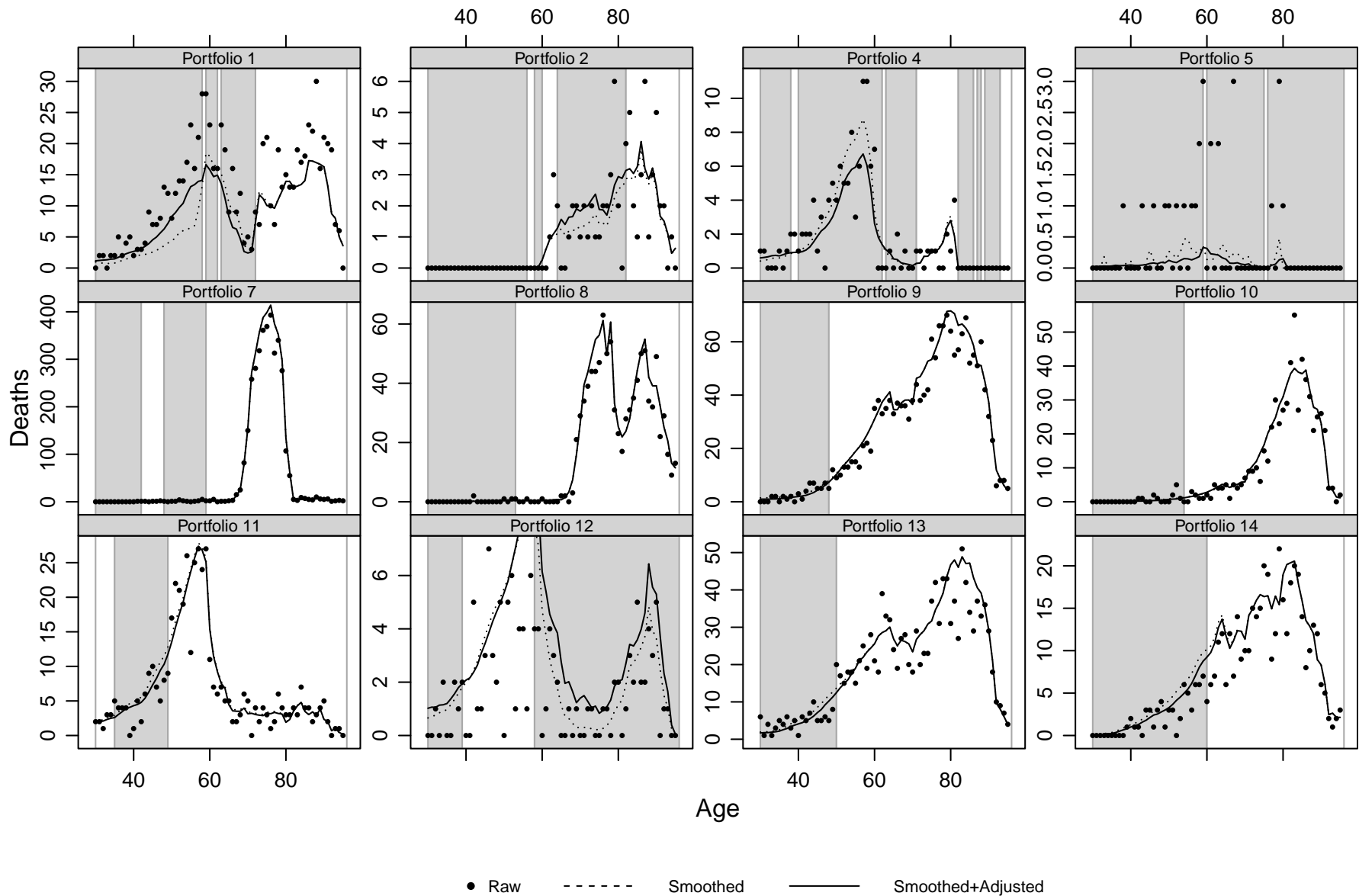
The visual inspection of the credibility based mortality curve shows that the regularity is preserved avoiding the limitation of the sole smoothing procedure discussed above. For some portfolios, such as portfolio P12, the regularization based on the credibility attached to each age level enhance the prediction of the future mortality. Indeed, the smoothed mortality based on past observations suggest a local distortion of the curve for ages ranging from 60 to 80. This particular feature is however not observed in the mortality curve for the year 2011 and thus the credibility based curve has a better fitting. This can also be observed in Figure 4, where the predicted deaths using  $\widetilde{\varphi}^i$ , for  $i = 12$ , are (visually) more in line with the observations. The same conclusions, in the grayed area, can be drawn for the other portfolios.

**4.4 Proximity Between the Observations and the Model.** Besides the visual inspection of the proposed adjustment, and in order to understand the impact of the the latter, we will use some known statistics to quantify the proximity between the observations and the outputs of the two curves considered in Figure 3 and Figure 4. We assess the overall deviation with the observed mortality by comparing criteria measuring the distance between the observations and the models with the  $\chi^2$ , i.e. Forfar et al. (1988b), the mean average percentage error (MAPE), see for instance



**Figure 3:** The next-period mortality rate based on the graduation of mortality (solid line) as well as its credibility based adjustment (dashed line) based on the period 2007 – 2010. Both predictions are compared to the observed mortality rate over the period 2011 (black dots).





**Figure 4:** The next-period deaths prediction based on the graduation of mortality (solid line) as well as its credibility based adjustment (dashed line) based on the period 2007 – 2010. Both predictions are compared to the observed deaths over the year 2011 (black dots).

Felipe et al. (2002); as well as the standardized mortality ratio (SMR), i.e. the ratio of deaths observed to those predicted. In addition, we find useful to use the SMR test proposed in Liddell (1984) and the likelihood ratio test to asses of significance of the SMR and to its interval estimation. The tests and quantities summarizing the proximity between the observations and the model, for each portfolio  $i$  at calendar year  $t = 2011$ , are described in the following.

(i) The  $\chi_i^2$  allows to measure the quality of the fit of the model. It writes,

$$\chi_i^2 = \sum_{x=x_1}^{x_n} \frac{(D_{x,t}^i - L_{x,t}^i \hat{q}_x^i)^2}{L_{x,t}^i \hat{q}_x^i (1 - \hat{q}_x^i)}.$$

(ii) The MAPE is the average of the absolute values of the deviations from the observations,

$$\text{MAPE}^i = \frac{\sum_{x=x_1}^{x_n} |(D_{x,t}^i/L_{x,t}^i - \hat{q}_x^i)/(D_{x,t}^i/L_{x,t}^i)|}{\sum_{x=x_1}^{x_n} D_{x,t}^i} \times 100.$$

(iii) The SMR is computed as the ratio between the observed and fitted number of deaths:

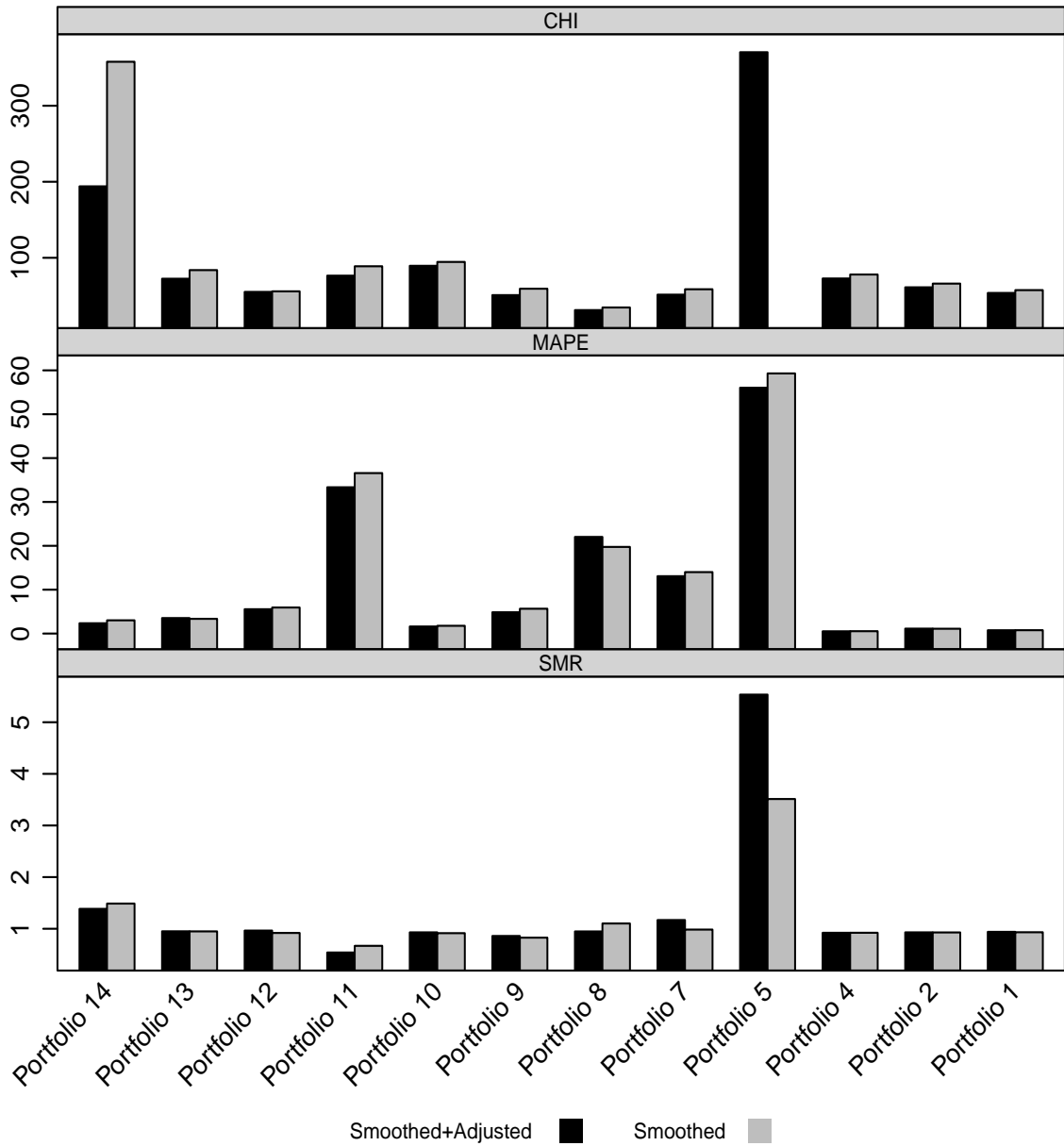
$$\text{SMR}^i = \frac{\sum_{x=x_1}^{x_n} D_{x,t}^i}{\sum_{x=x_1}^{x_n} L_{x,t}^i \hat{q}_x^i}.$$

Hence, if  $\text{SMR}^i > 1$ , the fitted deaths are under-estimated and vice-versa if  $\text{SMR} < 1$ . Note that we can consider the SMR as a global criterion which does not take the age structure into account, compared to the  $\chi^2$  and the MAPE for instance.

Figure 5 summarizes the above mentioned tests and quantities giving the overall deviation between the observations and the adjustment analysis for the portfolios P1 to P14 (except 3 and 6 which does not contain observations of year 2011) obtained by the smoothing approach together with the credibility adjustment procedure.

When looking at criteria and quantities which take the age structure of the error into account, the credibility approach has an important benefit compared to the sole graduated curve. The quality of the fit increases, sometimes drastically, i.e. portfolio P1, in terms of having the minimum  $\chi^2$  and MAPE values, see also the last panels of Table 3. Also, the credibility adjustment exhibits the highest  $p$ -value for the likelihood ratio test. Even when we consider a global indicator of the quality of the fit such as the SMR which does not take the age structure into account, the proposed procedure seems to perform better than the graduated curve. However, notice that the impact of adjustment is smaller when the portfolios are quiet big. This is already remarkable when one operated visual checks as mentioned earlier.

**4.5 Comparison with Classical Approaches.** Here, we wish to compare our model to the Hardy and Panjer (1998) and Poisson-Gamma credibility analysis applied to our mortality datasets. Moreover, we intend to compare our results to a similar approach introduced in Salhi et al. (2016), where the graduation of mortality is based on a parametric method, i.e. Makeham law, and the credibility theory is used to adjust the latter with incoming new data experience. The first two approaches focus on the actual to expected mortality ratio, in aggregate level, as a key observation. The adjustment is directly applied to this quantity. Specifically, the a priori expected number of deaths for each portfolio is updated at each period given the credibility weight on the observations coming from this portfolio and the one computed on the basis of the other portfolios. Table 3 presents the tests and quantities summarizing the overall deviation between the observations for



**Figure 5:** Distance between the observations and the models (top panel) together with their mean average percentage error between (middle panel) and the standardized mortality ratio (bottom panel).

these different approaches in comparison the one exposed in this paper.

We first note that the Hardy-Panjer and Poisson-Gamma approaches produce relatively similar graduations as the tests suggest sensibly similar outputs. However, we notice some differences with the Makeham credibility model which displays more favorable results. This is already outlined in [Salhi et al. \(2016\)](#) and this may be explained by the age-specific adjustment but thanks to the structural feature added by the Makeham parametric model. When it comes to the tests that are sensitive to the age structure, we notice that the credibility-based Makeham model offers a outstanding adjustment as the tests are favorable compared to the Hardy-Panjer and Poisson-Gamma approaches.

## 5 Concluding Remarks

In this paper we proposed a methodology to adjust the graduated mortality table that uses an adaptive smoothing procedure based on the local likelihood. The adjustment is based on the credibility weighting technique of the smoothed curves and a reference. Our approach takes into account the age specific heterogeneity that may arise in real world datasets. We thus consider updating the mortality for each age based on the upcoming past information from the same age but also the neighboring ages. The inclusion of the neighboring ages is crucial as the particular smoothing procedure used in this paper add a dependency between the single ages. Based on classical results on the inference of the smoothing procedure we derived the closed form formulas needed to adjust the mortality.

The proposed methodology is shown to outperform compared to the classical credibility approaches that does not take into account the age structure of the portfolio. This is in line with the recent work in this field as mentioned by [Salhi et al. \(2016\)](#). Even when the age structure is accounted for the methodology developed in this paper has an important benefit. This is mainly due to the underlying curve built using an adaptive procedure compared to the parametric model considered in [Salhi et al. \(2016\)](#).

We should note that the model proposed can be investigated in order to quantify mathematically the errors induced in the assessment of the next-period mortality curve. This amend to consider the uncertainty stemming from the estimation of the different variables used in the updating procedure. There are also several practical issues we do not address here such as the impact on the pricing of life insurance contracts. These are open questions that we openly acknowledge and leave for future research.

## References

- AALEN, O. O., “Heterogeneity in survival analysis,” *Statistics in medicine* 7 (1988), 1121–1137.
- BÜHLMANN, H. AND A. GISLER, *A course in credibility theory and its applications* (Springer, 2005).
- COPAS, J. AND S. HABERMAN, “Non-parametric graduation using kernel methods,” *Journal of the Institute of Actuaries* 110 (1983), 135–156.
- CURRIE, I. D., “Fitting models of mortality with generalized linear and non-linear models.,” Technical Report, 2013.
- DEBÓN, A., F. MONTES AND R. SALA, “A comparison of nonparametric methods in the graduation of mortality: Application to Data from the Valencia Region (Spain),” *International statistical Review* 74 (2006), 215–233.
- DELWARDE, A. AND M. DENUIT, *Construction de tables de mortalité périodiques et prospectives*, Assurance Audit Actuariat (Economica, 2005).

- DELWARDE, A., D. KACHKHIDZE, L. OLIE AND M. DENUIT, “Modèles linéaires et additifs généralisés, maximum de vraisemblance local et méthodes relationnelles en assurance sur la vie,” *Bulletin Français d’Actuariat* 6 (2004), 77–102.
- FELIPE, A., M. GUILLÉN AND A. PÉREZ-MARÍN, “Recent mortality trends in the Spanish population,” *British Actuarial Journal* 8 (2002), 757–786.
- FORFAR, D., J. MCCUTCHEON AND A. WILKIE, “On graduation by mathematical formula,” *Journal of the Institute of Actuaries* 115 (1988a), 1–149.
- , “On graduation by mathematical formula,” *Journal of the Institute of Actuaries* 115 (1988b), 1–459.
- GUSTAFSSON, J., J. P. NIELSEN, P. PRITCHARD AND D. ROBERTS, “Quantifying operational risk guided by kernel smoothing and continuous credibility,” *The ICAI Journal of Financial Risk Management* 3 (2006), 23–47.
- HARDY, M. R. AND H. H. PANJER, “A credibility approach to mortality risk,” *ASTIN Bulletin* 28 (1998), 269–283.
- HOUGAARD, P., “Life table methods for heterogeneous populations: distributions describing the heterogeneity,” *Biometrika* 71 (1984), 75–83.
- LIDDELL, F. D. K., “Simple exact analysis of the standardised mortality ratio,” *Journal of Epidemiology and Community Health* 38 (1984), 85–88.
- LOADER, C., *Local regression and likelihood* (Springer Science & Business Media, 2006).
- NIELSEN, J. P. AND B. SANDQVIST, “Credibility weighted hazard estimation,” *Astin Bulletin* 30 (2000), 405–418.
- OLIVIERI, A., “Heterogeneity in survival models. Applications to pensions and life annuities,” *Belgian Actuarial Bulletin* 6 (2006), 23–39.
- RAMLAU-HANSEN, H., “The choice of a kernel function in the graduation of counting process intensities,” *Scandinavian Actuarial Journal* 1983 (1983), 165–182.
- SALHI, Y., P.-E. THÉRON AND J. TOMAS, “A credibility approach of the Makeham mortality law,” *European Actuarial Journal* 6 (2016), 61–96.
- TIBSHIRANI, R. AND T. HASTIE, “Local likelihood estimation,” *Journal of the American Statistical Association* 82 (1987), 559–567.
- TOMAS, J., “A local likelihood approach to univariate graduation of mortality,” *Bulletin Français d’Actuariat* 11 (2011), 105–153.
- , “Univariate graduation of mortality by local polynomial regression,” *Bulletin Français d’Actuariat* 12 (2012), 5–58.
- TOMAS, J. AND F. PLANCHET, “Multidimensional smoothing by adaptive local kernel-weighted log-likelihood with application to long-term care insurance,” *Insurance: Mathematics & Economics* 52 (2013a), 573–589.
- , “Construction d’une table de mortalité par positionnement: mode d’emploi,” Note de travail III291-15 v1.0, Institut des Actuaire, 2013b.
- , “Construction et validation des références de mortalité de place,” Technical report III291-11 v1.3, Institut des Actuaire, 2013c.
- VAUPEL, J. W., K. G. MANTON AND E. STALLARD, “The impact of heterogeneity in individual frailty on the dynamics of mortality,” *Demography* 16 (1979), 439–454.
- WAND, M. P. AND M. C. JONES, *Kernel smoothing* (Crc Press, 1994).

## A Additional Tables

**Table 3:** Tests and quantities summarizing the deviation between the observations and the model

		Hardy-Panjer	Poisson-Gamma	Makeham-Credibility	Smoothed	Smoothed+Adj.
$\chi^2$	Portfolio 1	1901.240	1928.680	259.400	357.870	193.967
MAPE (%)		102.660	102.000	32.870	3.018	2.349
SMR		1.737	1.756	1.126	1.487	1.385
$\chi^2$	Portfolio 2	34.890	33.640	30.940	37.612	31.166
MAPE (%)		48.030	49.120	53.990	20.119	20.842
SMR		1.037	1.002	0.905	1.102	0.948
$\chi^2$	Portfolio 4	130.120	132.890	79.321	58.615	51.515
MAPE (%)		95.390	92.490	44.880	14.006	13.078
SMR		0.826	0.853	1.405	0.984	1.168
$\chi^2$	Portfolio 5	473.680	573.940	348.180	NA	370.401
MAPE (%)		85.660	88.040	90.420	59.296	56.038
SMR		2.857	3.424	5.021	3.513	5.534
$\chi^2$	Portfolio 7	221.640	223.560	195.000	77.997	72.795
MAPE (%)		135.390	135.710	37.250	0.534	0.509
SMR		0.846	0.844	0.823	0.922	0.922
$\chi^2$	Portfolio 8	2575.630	2583.900	2414.250	66.033	61.174
MAPE (%)		323.780	324.610	263.210	1.100	1.122
SMR		0.232	0.231	0.243	0.928	0.930
$\chi^2$	Portfolio 9	1572.530	1573.970	1502.870	57.461	53.735
MAPE (%)		368.080	368.290	125.640	0.764	0.755
SMR		0.423	0.423	0.419	0.932	0.940
$\chi^2$	Portfolio 10	115.820	116.470	97.880	83.790	72.448
MAPE (%)		89.680	91.030	46.140	3.356	3.530
SMR		0.871	0.862	0.960	0.948	0.950
$\chi^2$	Portfolio 11	415.320	417.530	76.480	55.888	55.127
MAPE (%)		152.870	151.690	46.970	5.934	5.548
SMR		0.829	0.837	1.018	0.918	0.964
$\chi^2$	Portfolio 12	130.050	129.230	90.740	88.836	76.459
MAPE (%)		110.540	107.220	95.270	36.577	33.344
SMR		0.598	0.619	0.543	0.669	0.539
$\chi^2$	Portfolio 13	351.560	351.360	263.550	94.570	89.428
MAPE (%)		180.910	180.610	54.620	1.765	1.608
SMR		0.839	0.840	0.832	0.914	0.930
$\chi^2$	Portfolio 14	227.860	227.950	85.920	59.317	50.885
MAPE (%)		159.740	160.600	53.530	5.659	4.852
SMR		0.792	0.788	0.939	0.827	0.860