Credibility Adjustment of the Lee-Carter Longevity Model for Multiple Populations

Jean-Baptiste Coulomb, Yahia Salhi, Pierre-Emmanuel Thérond

Credibility Adjustment of the Lee-Carter Longevity Model for Multiple Populations

Jean-Baptiste Coulomb,[1] Yahia Salhi[2] and Pierre-E. Thérond[2,3]

[1]PSL Research University, Université Paris Dauphine, Paris, France

[2]ISFA, Univ Lyon, UCBL, LSAF EA2429, F-69007, Lyon, France

[3]SeaBird, 15-17 Rue de Bucarest 75008, Paris, France

April 18, 2020

Abstract

In this paper we are interested in applying Bühlmann-Straub credibility for a multiple population model, in particular, in presence of small sized populations. We introduce a parsimonious extension of the classic Lee and Carter model in a similar manner as the joint-κ model, which does impose a common factor to the considered populations. Hence, we propose an adjustment procedure based on the Bühlmann-Straub linear credibility into a Lee-Carter model in order to take into account the heterogeneity among the populations while allowing for learning effect for each population. The updating mechanism is based on the credibility estimations of future mortality rates depending on past observations through a recursive credibility formula. By doing so, the forecasts weight the importance of the information stemming from the single population while taking into account the neighboring populations. The proposed methodology is applied to real-world datasets and comparisons to classic mortality models is proposed.

Keywords: Longevity; Mortality; Small Populations; Multi-Population; Credibility; Prediction

1 Introduction

Predicting future mortality patterns has been a great concern for a long time for demographers and actuaries in life insurance companies and pension funds. Today, insurers often rely on regulatory life tables to make projections, but these are sometimes too conservative. Therefore, this may lead to both an increase of *Best Estimate* technical provisions and thus a decrease of *Basic Own-Funds*, and an increase of the base figure used for calculating the solvency capital in charge for longevity risk. That is why, considering an adequate and coherent assumption on the future behaviour of mortality is of paramount importance. The European prudential regulation, Solvency II, has emphasized the need of mortality and life tables that best capture and reflect the experienced mortality in order to adequately quantify the underlying risk, see Salhi and Thérond (2018) and Barrieu et al. (2012) for more details.

Various approaches have been introduced in the actuarial literature to project mortality such as the seminal Lee and Carter's model, see Lee and Carter (1992), and the CBD models, see Cairns et al. (2006) and Cairns et al. (2011b). These propose a factor-based framework in order to decompose the mortality patterns into age, period and cohort effects. Recently, some alternative approaches have also been introduced, e.g. Doukhan et al. (2017, 2020) and Hainaut (2018). However, in life insurance, the size of the considered populations and the heterogeneity of the guarantees for the same underlying risk makes difficult the creation of such mortality tables based on the sole experience of each policy, which may induce significant impacts on the technical reserves if the tables have to be updated more frequently. This was the case, for instance, when the French prospective life tables were updated in 2007; replacing the previous set of tables from 1993. The resulting disparities between the 1993 prospective tables and observed longevity caused French insurers to sharply increase their reserves by an average of 8%.

Therefore, the goal is to model mortality rates specific to sub-populations with particular characteristics (population of a small country or region, individuals with a specific disease, insurance portfolio, sectorial pension funds or mortality by cause of death). Traditionally, relational models are used to overcome this issue, see Delwarde et al. (2004). By doing so, we are left with a bias known as basis risk, the mortality between the mortality in an individual portfolio and that of the national population - a result of selection effects, see Salhi and Loisel (2017). Also, these techniques, in the vein of the Cox (1992) proportional hazard framework, are static, and any discrepancy between the two sets of mortality figures represents a significant risk for the evaluation of the future mortality of the sub-population. Some recent literature looked at the modelling in the presence of sufficiently large sub-populations datasets, see Cairns et al. (2011a) and Coughlan et al. (2011). However, when it comes to the study of the mortality at a single portfolio or small sized populations, some specific issues arise. In fact, very few deaths are observable at some ages making it difficult to assess some of the parameters involved in the classic models. Thus, it is very important to be able to overcome these issues by proposing a well-suited methodology that adapts to these particular stylized facts. Recent literature, see e.g. Salhi et al. (2016) and Salhi and Thérond (2018), propose a framework to update the assumptions on the biometric risk of small populations using the credibility theory. The first methodology consists in using a parametric approach for graduating mortality with a Makeham mortality law. It is shown that a credibility approach and adjusting periodically the mortality level, the ability to predict death forecast is improved for population of small sizes, and that using a Makeham mortality law is better than using a Poisson-Gamma model. However, the proposed methodology is more adapted to mortality risk for which the assumptions may be adjusted on a yearly basis. On the other hand, the longevity risk assessment and modelling using the credibility theory become increasingly important over recent years. For instance, Schinzinger et al. (2016) proposed an evolutionary credibility model, using the Lee and Carter framework, that describes the joint dynamics of mortality through time in several populations. Such an approach is applied to large populations with mortality patterns that are usually much more regular and easier to accommodate using the Lee and Carter's model. In this paper, the adopted methodology roots on the recent advancement of multiple population modelling. In contrast to single population modeling, multi-population models are based on the assumption that mortality experience of one population may contain useful information for predicting mortality of another population. By incorporating this information, we allow for one population to learn from the others and thus increase the potential to improve mortality prediction of individual populations. As noted by Li and Lee (2005): "mortality patterns and trajectories in closely related populations are likely to be similar in some respects, and differences are unlikely to increase in the long run". The authors proposed an extension of the Li and Lee (2005) method to coherently forecast mortality for a group of populations. Their approach assumes that close populations in a given group must have the same drift driving the mortality pattern as well as the age-specific sensitivity. Their modelling approach assumes then an additional common factor in the same line as in the initial Lee and Carter's model. By doing so, it is possible to ensure that mortality forecasts in different populations are coherent, i.e., projected mortality differentials of different populations should not diverge in the long run. Formally, we will suppose that the mortality profile for each population follows a common factor model as introduced by Li and Lee (2005) and referred to as the Joint-κ model, see Li and Hardy (2011), which amends to say that the evolution of mortality follows similar patterns for the different populations. Hence, we aim at using the available information stemming from the other populations to sequentially adjust the mortality forecasts. This will implicitly invoke the use of credibility theory in the same line as Salhi et al. (2016) and Salhi and Thérond (2018). Similarly, Tsai and Lin (2017b) introduced the Bülmann credibility into the classic stochastic mortality models, with an objective of improving their forecasting performance. Also, Tsai and Lin (2017a) used a similar approach in order to enhance the predictive power of the classic multiple population models. In their work the information stemming from each population plays the same role in the sense that the weight assigned to each population does not depend neither on the size nor the credibility of the information itself. However, the weight is of paramount importance. It reflects the importance of the information flow over years and implicitly links the variability of the estimation of the parameters to the size of the underlying population: very small portfolios are subject to larger estimation variability and vice versa. Taking into account the weight is critical in assigning a credibility factor to the information stemming from each population. Therefore, in this paper, the main objective is to allow for learning effect among the populations and thus expect the forecast to weight the importance of the information stemming from the single population while taking into account the neighboring populations.

The paper is organized as follows. In Section 2, we recall the classic mortality approaches starting with the seminal Lee and Carter's model and its variant. The multiple population modelling approach is introduced. Then, we motivate the modelling framework which is an extension of the Lee and Carter's model. We discuss the limit of such a model, in particular, when it comes to modelling small populations. In Section 3, we propose a credibility-based adjustment for the common-factor model, also referred to as the joint κ model. In this section, we propose the estimation of the updating factors allowing to learn from the neighboring populations. Finally, in Section 4, we detail the updating mechanism and propose an application to real-world datasets.

2 Mortality Models and Multiple Population Extension

2.1 Lee and Carter's Factor-Based Modeling Approach. There is an abundant literature that propose to predict the evolution of human mortality using stochastic models. In classic dynamic mortality models, death probabilities (or mortality intensities) are represented as functions of age, period (calendar year) and cohort (year of birth) parameters, e.g. Lee and Carter (1992), Cairns et al. (2006) and Renshaw and Haberman (2003) among others. Most commonly used mortality models focus on mortality rates time series, starting with the famous Lee-Carter model Lee and Carter (1992), widely used by insurance practitioners and demographers. This model describes the intensity of mortality $m_{x,t}$ at age x and time t using three parameters, namely α_x, β_x and κ_t as follows:

$$\log(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_t(x), \quad \epsilon_t(x) \sim \mathcal{N}(0, \sigma), \tag{2.1}$$

where α_x gives the average level of mortality at each age over time; the time varying component κ_t is the general speed of mortality improvement over time and β_x is an age-specific component that characterizes the sensitivity to κ_t at different ages ; the β_x also describes (on a logarithmic scale) the deviance of the mortality from the mean behaviour, κ_t. The error term $\epsilon_t(x)$ captures the remaining variations.

In general, insurance companies and pension funds are facing the heterogeneous evolution of policyholders mortality making it difficult to apply directly the classic models to their portfolios. In fact, the Lee and Carter model and more generally the class of factor-based models described above is applied to large datasets, and thus are most suited to model mortality rates for an underlying national population. However, as noted the final goal is to model mortality rates specific to sub-populations with particular characteristics (population of a small country or region, individuals with a specific disease, insurance portfolio or sectorial pension funds). Traditionally, relational models are used to overcome this issue, see Delwarde et al. (2004). By doing so, we are left with a bias known as basis risk, the mortality between the mortality in an individual portfolio and that of the national population - a result of selection effects, see Salhi and Loisel (2017). Also, these techniques, in the vein of the Cox (1992) proportional hazard framework, are static, and any discrepancy between the two sets of mortality figures represents a significant risk for the evaluation of the future mortality of the sub-population. Some recent literature looked at the modelling in particular in the presence of sufficiently large sub-populations datasets, see Cairns et al. (2011a) and Coughlan et al. (2011). However, when it comes to the study of the mortality at a single portfolio or small sized populations, some specific issues arise. In fact, very few deaths are observable at some ages making it difficult to assess some of the parameters involved in the model described above. For instance, when it comes to the estimation of the age-specific mortality profile

β_x, one needs sufficiently large observations at the age level in order to evaluate the sensitivity of the mortality decline at each age. Secondly, available age-specific mortality statistics lacks of deepness, making it harder to isolate properly the time trend parameter κ_t, see Salhi et al. (2016) and Salhi and Thérond (2018). Therefore, it is very important to be able to overcome these issues by proposing a well-suited methodology that adapts to these particular stylized facts. Recent literature, see e.g. Salhi et al. (2016) and Salhi and Thérond (2018), propose a framework to update the assumptions on the biometric risk of small populations using the credibility theory. The first methodology consists in using a parametric approach for graduating mortality with a Makeham mortality law. It is shown that a credibility approach and adjusting periodically the mortality level, the ability to predict death forecast is improved for population of small sizes, and that using a Makeham mortality law is better than using a Poisson-Gamma model. However, the proposed methodology is more adapted to mortality risk for which the assumptions may be adjusted on a yearly basis. On the other hand, the longevity risk assessment and modelling using the credibility theory become increasingly important over recent years. For instance, Schinzinger et al. (2016) proposed an evolutionary credibility model, using the Lee and Carter framework, that describes the joint dynamics of mortality through time in several populations. Formally, the same age effect β_x was used for all groups to avoid long-run divergence in gender-specific forecast which is an important criterion for modelling mortality of closely related populations. However, such an approach is applied to large populations with mortality patterns that are usually much more regular and easier to accommodate using the Lee and Carter's model. Similarly, Tsai and Lin (2017b) introduced the Bülmann credibility into the classic stochastic mortality models, with an objective of improving their forecasting performance.

In this paper, the adopted methodology roots on the recent advancement of multiple population modelling. In contrast to single population modeling, multi-population models are based on the assumption that mortality experience of one population may contain useful information for predicting mortality of another population. By incorporating this information, we allow for one population to learn from the others and thus increase the potential to improve mortality prediction of individual populations. As noted by Li and Lee (2005): "mortality patterns and trajectories in closely related populations are likely to be similar in some respects, and differences are unlikely to increase in the long run". The authors proposed an extension of the Li and Lee (2005) method to coherently forecast mortality for a group of populations. Their approach assumes that close populations in a given group must have the same drift κ_t driving the mortality pattern as well as the age-specific sensitivity β_x, see Equation (2.1). Their modelling approach assumes then an additional common factor in the same line as in the initial Lee and Carter's model. By doing so, it is possible to ensure that mortality forecasts in different populations are coherent, i.e., projected mortality differentials of different populations should not diverge in the long run.

2.2 Extended Lee and Carter's Model. When dealing with longevity risk, we are generally facing various sources of uncertainty. Usually actuaries do consider two main components which are mainly related to idiosyncratic and systematic aspects of such a risk. First, there is a volatility component, whether intrinsic or due to small sampling sized populations. Secondly, we are dealing with the *level risk*, which is related to the risk of mi-specifying the initial level or mortality. This can be, for instance, explained by the uncertainty surrounding the estimation of the parameter α_x

in model (2.1). Finally, there is the trend risk which amounts to underestimation of the mortality decline over time. For the Lee and Carter's model this corresponds to the uncertainty of estimating the multiplicative parameters β_x and κ_t. As noted before, these different sources of risk are highly amplified for small populations, because the volatility may be higher and the trend as well as the level may differ from those of a bigger population. However, it is better not to simply use a baseline mortality level and a baseline mortality trend as the goal is to better assess the risks associated to the small population considered.

First, regarding the level and as soon as it is not the same as the baseline, it will be determined using the classic Lee-Carter model. $\cdots$

$$\log(m^i_{x,t}) = \alpha^i_x + \beta^i_x \kappa^i_t + \epsilon^i_{x,t} \ , \quad x = \underline{x}, \ldots, \overline{x}, \quad t = t_0, \ldots, T, \quad i = 1, \ldots, I \ . \tag{2.2}$$

where $\epsilon^i_{x,t} \sim \mathcal{N}\left(0, (\sigma^i_x)^2\right)$ and two constraints: $\sum_{x=\underline{x}}^{\overline{x}} \beta^i_x = 1$ and $\sum_{t=t_0}^{T} \kappa^i_t = 0$ for all $i = 1, \ldots, I$. As briefly discussed in the above section, using such a model poses some challenges. In fact, when dealing with small populations (countries or insured populations), at least two questions need to be answered. First, the underlying population are generally of small size, so very few deaths are observable at some ages and may cause high fluctuations for the death rate $m^i_{x,t}$, which will impact the parameters of the model in (2.2). As noted by Yue et al. (2019), "the biased estimates in the case of small populations are probably the main reason why many recent studies focus on modifying mortality models for small populations". Secondly, the available age-specific mortality statistics for such populations lack of deepness. This will then arise question on the quality of results than can be expected to derive through the Lee and Carter's model, see Li et al. (2004). Indeed, this makes difficult to isolate a possible time trend as it may be captured by κ^i_t. A natural approach to cope with this undesired effect is to include the mortality data from populations with similar mortality profiles. For instance, Li and Lee (2005) proposed a coherent extension of the Lee and Carter's model that aims at reducing the aforementioned estimation's errors by linking the mortality data from populations with similar mortality improvements. The model presents an augmented common factor model for a group of populations. It imposes a common mortality change by age but allows each population to develop its own age pattern and level of mortality. Formally, the augmented common factor model writes

$$\log(m^i_{x,t}) = \alpha^i_x + \beta^i_x \kappa^i_t + \beta^\bullet_x \kappa^\bullet_t + \epsilon^i_{x,t} \ , \quad x = \underline{x}, \ldots, \overline{x}, \quad t = t_0, \ldots, T, \quad i = 1, \ldots, I \ , \tag{2.3}$$

where, here and in the sequel, the subscript "$\bullet$" refers to summation over the corresponding index. In the above model the common parameters $\beta^\bullet_x$ and $\kappa^\bullet_t$ are estimated from the combined data for all populations. The population specific parameters α^i_x, β^i_x and κ^i_t have the same interpretation as in the initial Lee and Carter's model.

In their work, Li and Lee (2005) discussed the importance of including the population specific component $\beta^i_x \kappa^i_t$ in model (2.3). The so-called explanation ratio, which measures the goodness-of-fit of a model for the ith population, is used to assess the importance of the specific component. The outcomes of their goodness-of-fit analyses point out the relative importance of such a specific component only in few cases. In fact, using mortality of 11 countries it is shown that the explanation ratio is high enough, suggesting that there indeed exists a common trend for these countries, and that the common factor, without the specific component, model captures this trend quite well. Moreover, as discussed by Enchev et al. (2017), the augmented model (2.3) incorporates a significant number of parameters, which result in a considerable computing time and may also alter the forecasts robustness. Various modifications with a simplified structure and assumptions of this model were analyzed. All these models have identifiability problems that have been addressed by applying exact and quasi identifiability constraints. Although the Li and Lee's model in Equation (2.3) is showed to fit the data quite well, it exhibited robustness problems along with associated problems with slow convergence. Enchev et al. (2017) pointed to a need for caution in the use of that model and, potentially, for the introduction of some quasi identifiability or other constraints. It is finally showed that a model with a common age effect, see Kleinow (2015), and an adjusted time-trend in the form of $\beta_x^\bullet(\kappa_t^\bullet - \kappa_t^i)$ performs fairly well on the considered mortality. This reinforces the idea that the common factor model without population-specific additional adjustment could be a good basis for modelling a group of population. However, we should be able to adjust for potential local divergence of mortality in the short term. This will be the starting point of the construction of our model. We should also note that other authors have also carried out further work, see Li and Hardy (2011) and Danesi et al. (2015), which draw similar conclusions on the common factor approach.

Therefore, in the sequel and based on these discussions, we will suppose that the mortality profile for each population follows a common factor model as introduced by Li and Lee (2005) and referred to as the Joint-κ model, see Li and Hardy (2011), and thus the following assumption holds.

(H1) For each $i = 1, \ldots, I$ we let $\kappa_t^i = \kappa_t^\bullet$,

which amends to say that the evolution of mortality follows similar patterns for the different populations. Mathematically, the model can be expressed as follows:

$$\log(m_{x,t}^i) = \alpha_x^i + \beta_x^i \kappa_t^\bullet + \epsilon_{x,t}^i \ , \quad x = \underline{x}, \ldots, \overline{x}, \quad t = t_0, \ldots, T, \quad i = 1, \ldots, I \ . \tag{2.4}$$

Similarly to the previous approaches, we will use a two-step approach to estimate this model. First, $\kappa_t^\bullet$ should be obtained from applying the ordinary Lee and Carter's method to the whole group, i.e. $m_{x,t}^\bullet$. The remaining population-specific parameters α_x^i and β_x^i can be estimated from the ordinary least-squares (OLS) regression. We can also obtain the parameter α_x^i by setting it to the average of $\log m_{x,t}^i$ and then regress $\log m_{x,t}^i - \alpha_x^i$ on $\kappa_t^\bullet$ without the constant term for each age x to get β_x^i.

2.3 Joint-κ Population-Specific Adjustment. The common factor model in Equation (2.4) offers a parsimonious and transparent way of assessing mortality differentials. However, it may be too stringent for some applications. In particular, when small populations are concerned, the estimation of some parameters would pose some problems. First, it has large number of parameters, even if these were reduced compared to the augmented common factor. In particular, the specific factor β_x^i can produce excessive divergences between projections for the different populations. As noted by Debón et al. (2011), the age-specific sensitivity parameter can behave erratically for small populations. This particular behavior is due to the lack of sufficiently large information on the mortality at the age level as well as the estimation bias of the initial mortality, see Salhi et al. (2016) and Salhi and Thérond (2018) for instance. In order to overcome this issue, we will introduce the following assumption.

(H2) For each $i = 1, \dots, I$, we assume that the ratio $\beta^i_x/\beta^\bullet_x$ is age-independent.

This hypothesis assumes the stability of the variability in the improvement rates for the considered population with respect to the aggregated population. Furthermore, in order to link this to the actuarial literature for multiple population modelling, we let $X^i = \beta^i_x/\beta^\bullet_x$ be the ratio considered in **(H2)**. Then, as with the joint-κ model, the interpretation of mortality differentials can be made easier under this assumption. Formally, we consider a re-parametrisation of Equation (2.4) in the form

$$\log(m^i_{x,t}) = \alpha^i_x + X^i\beta^\bullet_x\kappa^\bullet_t + \epsilon^i_{x,t}\ , \quad x = \underline{x}, \dots, \overline{x}, \quad t = t_0, \dots, T, \quad i = 1, \dots, I\ . \tag{2.5}$$

Thus, we can recognise the three-way decomposition of the mortality rates introduced by Russolillo et al. (2011). It gives a straightforward interpretation of mortality evolution among the populations and decompose the mortality evolution into three factors: age, time and population. The latter is an indicator of the improvement differentials among populations. In fact, a population with faster (slower) rate than the mortality of global population has a parameter $X^i > 1$ ($X^i < 1$) see Villegas and Haberman (2014) and Villegas et al. (2017). As argued in Carter and Lee (1992), this simple arrangement may enforce greater consistency and is a parsimonious way to model multiple populations. However, by doing so, Assumption **(H2)** implicitly assumes that the death rates of the considered populations are perfectly associated, an assumption with limited empirical underpins and most notably important consequences in risk management. Secondly, when considering small populations, mortality differentials captured by the population-factor X^i may change from an estimation period to another. In fact, the instability discussed in the previous sections will undoubtedly be recovered on this parameter making it vary (considerably) over time. Therefore, it is important to correctly capture this effect.

Therefore, and due to these different sources of uncertainty on the assessment of the differential mortality level X^i, among other things, we suppose that the true level X^i is known up to unobservable factor Θ_i. In fact, one may think of the I populations as a subset of the aggregate population and thus each population is characterized by a risk profile Θ_i. This can be seen as a random effect or heterogeneity characterizing the specific differential mortality for each population. This invokes the use of the credibility theory in order to assess the future evolution of each X^i, while taking into account the information stemming from the remaining populations.

3 Credibility Adjustment for Joint-κ Lee and Carter's Model

We suppose that the model in Equation (2.4) is estimated on each period $[t_0, t]$ and denote $X^i_t = \beta^i_x/\beta^\bullet_x$ the corresponding population-specific differential ratio. The time index for the parameter X^i_t relates to the estimation period. We are now at time T and observations $X^i_{t_0}, \dots, X^i_T$ are the available information at this time. The aim is then to predict the conditional future expected observations $\mathbb{E}[X^i_{T+k}|\Theta_i]$, with $k = 1, 2, \dots,$ for each population $i = 1, \dots, I$.

3.1 Assumptions and Credibility Framework. The conditional expected differentials, i.e. $\mathbb{E}[X^i_{T+k}|\Theta_i]$, is of paramount importance in predicting future evolution of mortality rates. The main objective is to allow for learning effect among the populations and thus expect the forecast to weight the importance of the information stemming from the single population while taking into account the neighboring populations. This implicitly invoke the use of the credibility theory, see e.g. Bühlmann and Straub (1970) who developed the theoretical foundation of modern credibility theory. This approach has been widely applied in non-life insurance and more recently in life insurance, see among others Hardy and Panjer (1998), Salhi et al. (2016), Salhi and Thérond (2018), Tsai and Lin (2017a,b) and Li and Lu (2018). The fundamental idea underlying the credibility approach is the construction of a yearly decrement of the differential level as weighted average of the aggregate level and the sample mean of the past observations. In other words, the objective is to propose an estimation for $\mathbb{E}[X^i_{T+K}]$ based on the population mortality experience balanced with the collective observations. However, in order to adequately constructed this estimator we shall need the following classic assumptions.

(H3) Conditionally to the risk profile $\Theta_i = \theta_i$, the random variables X^i_t, for $t = t_0, \cdots, T$, are i.i.d. with distribution given by G_{θ_i}.

(H4) The two first moments $\mu(\theta_i) = \int_0^{+\infty} x \; dG_{\theta_i}(x)$ and $\sigma^2(\theta_i) = \int_0^{+\infty} (x - \mu(\theta_i))^2 \; dG_{\theta_i}(x)$ are finite.

Assumption **(H3)** translates the dependency of the differential mortality over time. It is only captured by the risk profile Θ_i. Conditionally on the knowledge of this risk parameter, the successive differential levels of mortality with regard to the aggregate are independent. Finally, the last hypothesis **(H4)** assumes the existence of the conditional moments, which is a critical condition to ensure the convergence of the mortality rates over time. In fact, if any of the two first moments diverges, it would implicitly induce a divergence of the forecast and thus the divergence of mortality rates among the considered populations. Indeed, it is usually desirable that the forecasts do not diverge over time. This is referred to as the coherence property, see Li and Lee (2005), and ensures that the forecast maintain a given structural relationship and thus populations with similar mortality profiles do not diverge in the long run, see Salhi and Loisel (2017).

Hereafter, for each population i, we focus on the projection of X^i_{T+1}. We will detail the adjustment mechanism in the following sections for $k = 2, 3, \ldots$. From a more mathematical point of view, a collection of I random variables $(X^i\Theta_i)$ is considered. Population-specific observations are assumed to be an outcome of the random vector $(X^i_t)_{t=t_0,\ldots,T}$. The objective is to estimate the next period projection of the differential ratio for each portfolio i. More precisely, in view the available data up to time T, one aims to find the best estimate of $\mathbb{E}\left[X^i_{T+1}|\Theta_i\right]$, which is unknown. Here, we will look for the best linear predictor denoted $\mu(\Theta_i)$ in terms of the random vector $(X^i_t)_{t=t_0,\ldots,T}$. Let $\widehat{\mu}(\Theta_i)$ be this estimation. For this purpose and using the usual credibility setting, we shall make the following additional hypotheses.

(H5) For each $t = t_0, \ldots, T$ and $i = 1, \ldots, I$, $\mathbb{E}\left[X^i_t|\Theta_i\right] = \mu(\Theta_i)$ and $\mathbb{V}\mathrm{ar}\left[X^i_t|\Theta_i\right] = \sigma^2(\Theta_i)/\omega^i_t$ where ω^i_t is a predefined *weight*.

(H6) The pairs $(\Theta_i, X^i_t), (\Theta_k, X^k_t), k \neq i$ are independent and identically distributed.

Assumption **(H5)** implies that for each population i, the *true* relative ratio $\mu(\Theta_i)$ (conditionally on the knowledge of the risk profile Θ_i) does not change over time, and its variance given by Θ_i, $\mathbb{V}\mathrm{ar}\left[X_t^i|\Theta_i\right]$ changes in proportion to the relative importance of the portfolio. Here, the weight ω_t^i reflects the importance of the information flow stemming from population i over the year t. It implicitly links the variability of the estimation of the parameter X_t^i to the size of the underlying population: very small portfolios are subject to larger variability on the estimation of X^i and vice versa. Taking into account the weight is critical in assigning a credibility factor to the information coming from each population. In fact, as discussed earlier, for small sized populations the variance of the differential level X^i. For instance, in Tsai and Lin (2017b,a), it is assumed that the populations of interest have common (conditional) mean and variance. Although, the underlying model used the improvement rates, i.e. $\log m_{x,t}^i - \log m_{x,t-1}^i$, as a risk factor, it still shows some limits when it comes to the modelling of small sized populations. In fact, the authors only focus on the populations (male and females) of the United Kingdom and Japan.

Finally, Assumption **(H6)** means that the risk profiles are independent. The successive realizations of the relative ratio X^i, i.e. X_t^i for $t = t_0, \ldots, T$, for any portfolio are independent of each other except through the risk parameter Θ_i. Intuitively, Assumption **(H6)** implicitly suggests that populations are comparable as they share common characteristics and are related to the same reference population, but not entirely similar which induces the conditional independence. These assumptions imply the following results.

(i) The expected prediction of X_{T+1}^i unconditionally on the risk profile Θ_i is given by $\mathbb{E}\big[X_{T+1}^i\big] = \mathbb{E}\big[\widehat{X}_{T+1}(\Theta_i)\big] = 1$. In other words, in the absence of any information on the heterogeneity level on the parameter X_i, the best next-period prediction of the latter is set to 1. Alternatively, when it comes to the modelling of mortality rates, this means that the common factor model of Li and Lee (2005) in Equation (2.4) is best suited for such a population, i.e. $\mathbb{E}\left[\beta_x^i\right] = \beta_x^\bullet$.

(ii) Using the law of total variance, the dependence structure of population i's relative level with the associated risk factor over time is given as

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(X_l^i, X_t^i) &= \mathbb{C}\mathrm{ov}\left(\mathbb{E}\left[X_l^i(\Theta_i)\right], \mathbb{E}\left[X_t^i(\Theta_i)\right]\right) + \mathbb{E}\left[\mathbb{C}\mathrm{ov}(X_l^i, X_t^i|\Theta_i)\right], \\
&= \mathbb{V}\mathrm{ar}\left[\mu(\Theta_i)\right] + \mathbb{E}\left[\mathbb{C}\mathrm{ov}(X_l^i, X_t^i|\Theta_i)\right], \\
&= \begin{cases} \tau^2, & \text{if} \quad l \neq t, \\ \tau^2 + \dfrac{\sigma^2}{\omega_t^i}, & \text{if} \quad l = t, \end{cases}
\end{aligned} \tag{3.1}$$

for $l, t \in \{t_0, \ldots, T\}$, where $\mathbb{V}\mathrm{ar}\left[\mu(\Theta_i)\right] = \mathbb{V}\mathrm{ar}\left[\Theta_i\right] := \tau^2$ and $\mathbb{E}\left[\sigma^2(\Theta_i)\right] = \mathbb{E}\left[\Theta_i\right] := \sigma^2$.

3.2 Credibility Estimator and Adjustment for the Differential Levels. In order to estimate $\mu(\Theta_i)$ we use the Bühlmann and Straub's credibility approach. It consists in estimating the next-period best prediction as a projection in a relevant Hilbert space onto a linear subspace. In other words, the relative mortality improvement for portfolio i and year $T+1$ belongs to class of linear estimators in observations and has the form $c_0^i + c_{t_0}^i X_{t_0}^i + \cdots + c_T^i X_T^i$. Hence, the expected mortality differential $\mathbb{E}[X_{T+1}|\Theta_i] = \mu(\Theta_i)$ will be estimated by $\widehat{\mu}(\Theta_i)$ which is of the form:

$$\widehat{\mu}(\Theta_i) = \widehat{c}_0^i + \sum_{t=t_0}^{T} \widehat{c}_t^i X_t^i, \tag{3.2}$$

where the coefficients $\widehat{c}_t^i$ for $t = t_0, \ldots, T$ are those minimizing the mean squared errors criterion:

$$(\widehat{c}_t^i)_{t=t_0,\ldots,T} = \underset{(c_t^i)_{t=t_0,\ldots,T}}{\operatorname{argmin}} \left\{ \mathbb{E}\left[\left(\widehat{\mu}(\Theta_i) - c_0^i - \sum_{t=t_0}^{T} c_t^i X_t^i \right)^2 \right] \right\}, \tag{3.3}$$

where the expectation is over the joint distribution of (X^i, Θ_i). In view of Equation (3.1) and taking the derivatives of the above criterion with respect to the c_t^i's and equalizing to zero gives the following estimations

$$\widehat{c}_0^i = 1 - \frac{\tau^2 \overline{\omega}^i}{\sigma^2 + \tau^2 \overline{\omega}^i} \quad \text{and} \quad \widehat{c}_t^i = \frac{\tau^2 \omega_t^i}{\sigma^2 + \tau^2 \overline{\omega}^i} \quad \text{with} \quad \overline{\omega}^i = \sum_{t=t_0}^{T} \omega_t^i. \tag{3.4}$$

Using these estimations, we can derive the Bühlmann and Straub's credibility estimator of X_{T+1}^i by substituting (3.4) into (3.3):

$$\widehat{X}_{T+1}^i(\Theta_i) = Z_i \overline{X}^i + (1 - Z_i)\widehat{\mu}, \tag{3.5}$$

with

$$Z_i = \frac{\tau^2 \overline{\omega}^i}{\sigma^2 + \tau^2 \overline{\omega}^i}, \qquad \overline{X}^i = \frac{1}{\overline{\omega}^i} \sum_{t=t_0}^{T} \omega_t^i X_t^i \quad \text{and} \quad \widehat{\mu} = \frac{\sum_{i=1}^{I} \widehat{Z}_i \overline{X}^i}{\sum_{i=1}^{I} \widehat{Z}_i}. \tag{3.6}$$

Here, Z_i represents an individual weight on historical observations, is called the *credibility factor* for portfolio i and takes values in $[0, 1]$. For each portfolio i the larger the volume of historical data, the larger Z_i will be (close to 1) .

3.3 Estimators of the Structure Parameters. As the risk parameters Θ_i are assumed to be identically distributed, their moments are identical, see Assumption **(H6)**. Therefore, τ^2 and σ^2 are the same for all populations and measure the residual heterogeneity of the risk profiles and the pure randomness respectively. These parameters are the key determinants of the credibility estimator. Let us recall the definition of the structure parameters:

$$\sigma^2 = \mathbb{E}\left[\sigma^2(\Theta_i)\right] = \omega_t^i \mathbb{E}\left[\mathbb{V}\text{ar}\left[X_t^i|\Theta_i\right]\right] \quad \text{and} \quad \tau^2 = \mathbb{V}\text{ar}\left[\mathbb{E}\left[X_t^i|\Theta_i\right]\right]. \tag{3.7}$$

Following Bühlmann and Straub (1970), we propose the estimators $\widehat{\sigma}^2$ and $\widehat{\widehat{\tau}}^2$, based on the observations $(X_t^i)_{t=t_0,\cdots,T}$:

$$\widehat{\sigma}^2 = \frac{1}{I} \sum_{i=1}^{I} s_i^2, \quad \widehat{\widehat{\tau}}^2 = \frac{\overline{\omega}}{(\overline{\omega})^2 - \sum_{i=1}^{I} (\overline{\omega}^i)^2} \left\{ \sum_{i=1}^{I} \overline{\omega}^i (\overline{X}^i - \overline{X})^2 - (I-1)\widehat{\sigma}^2 \right\},$$

with

$$s_i^2 = \frac{1}{T-t_0}\sum_{t=t_0}^{T}\omega_t^i (X_t^i - \overline{X}^i)^2, \quad \overline{X} = \frac{1}{\overline{\omega}}\sum_{i=1}^{I}\overline{\omega}^i\overline{X}^i \quad \text{and} \quad \overline{\omega} = \sum_{i=1}^{I}\overline{\omega}^i.$$

These estimators are unbiased and consistent, see Bühlmann and Straub (1970). We may note that $\widehat{\widehat{\tau}}^2$ can be negative. In such case $\widehat{\widehat{\tau}}^2$ is set to 0 which means that there would be no difference between the risks. The trend $\kappa_t^\bullet$ will be the same for all the sub-populations, see Assumption **(H3)**, and thus the common factor model of Li and Lee (2005) in Equation (2.4) is used to predict future pattern of mortality.

The credibility formula can now be evaluated by replacing its structural parameters by their consistent estimators. Formally, the empirical credibility estimator for X_{T+1}^i is obtained from (3.1) by replacing the structural parameters σ^2 and τ^2 by their estimators $\widehat{\sigma}^2$ and $\widehat{\tau}^2$, which can be written as follows:

$$\widehat{\widehat{X^i}}_{T+1} = \widehat{Z}_i\overline{X}^i + (1-\widehat{Z}_i)\widehat{\mu}, \quad \text{with} \quad \widehat{Z}_i = \frac{\widehat{\tau}^2\overline{\omega}^i}{\widehat{\sigma}^2 + \widehat{\tau}^2\overline{\omega}^i}. \tag{3.8}$$

It thus follows, from Equation (2.5), that the forces of mortality can be successively updated as follows:

$$\widehat{m}_{x,T+1}^i = \exp\left(\widehat{\alpha}_x^i + \widehat{\beta}_x^\bullet \widehat{X}_{T+1}^i \widehat{\kappa}_{T+1}^\bullet\right) = \exp\left(\widehat{\alpha}_x^i + \widehat{\beta}_x^\bullet\left(\widehat{\mu} + \widehat{Z}_i(\overline{X}^i - \widehat{\mu})\right)\widehat{\kappa}_{T+1}^\bullet\right). \tag{3.9}$$

4 Numerical Analyses

4.1 Data. We perform our analysis based on data obtained from the Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)[1]. The typical data set consists of the numbers of deaths $D_{x,t}^i$ and the central exposure $E_{x,t}^i$. We wanted to work with dataset covering a longer period in such a way that an out-of-sample analysis can be carried out. In fact, a first time-frame will be used to calibrate the models as well as the credibility updating procedure and a second period is used to assess the accuracy of forecast. Hence, the age and time periods range considered is from 55 up to 95 years-old and from 1975 up to year 2014 (40 years in total). Table 1 specifies the countries that were chosen. They were explicitly selected to have different population sizes. These are all countries of the Humand Mortality Database for which the data are available over the desired period and age band.

4.2 Joint-κ Model Estimation. The model considered in Equation (2.4) can be estimated using the two-step procedure discussed in Subsection 2.2, which gives the parameters $\kappa_t^\bullet$ and $\beta_x^\bullet$ needed for our model. Recall that these are estimated using the usual procedure proposed by Lee and Carter and based on the singular value decomposition on the matrix $\log m_{x,t}^\bullet - \alpha_x^\bullet$, where $\alpha_x^\bullet$ is the average mortality (on logarithmic scale) over the considered period. Hence, the estimation of the level X_t^i for each period $[t_0, t]$, where $t = t_0, \ldots, T$, can be done by equating it to the ratio of the estimated parameters β_x^i and $\beta_x^\bullet$ over the same period. Here, we rather advocate the use of

[1]The data is available at `www.mortality.org` (data downloaded on September 2019)

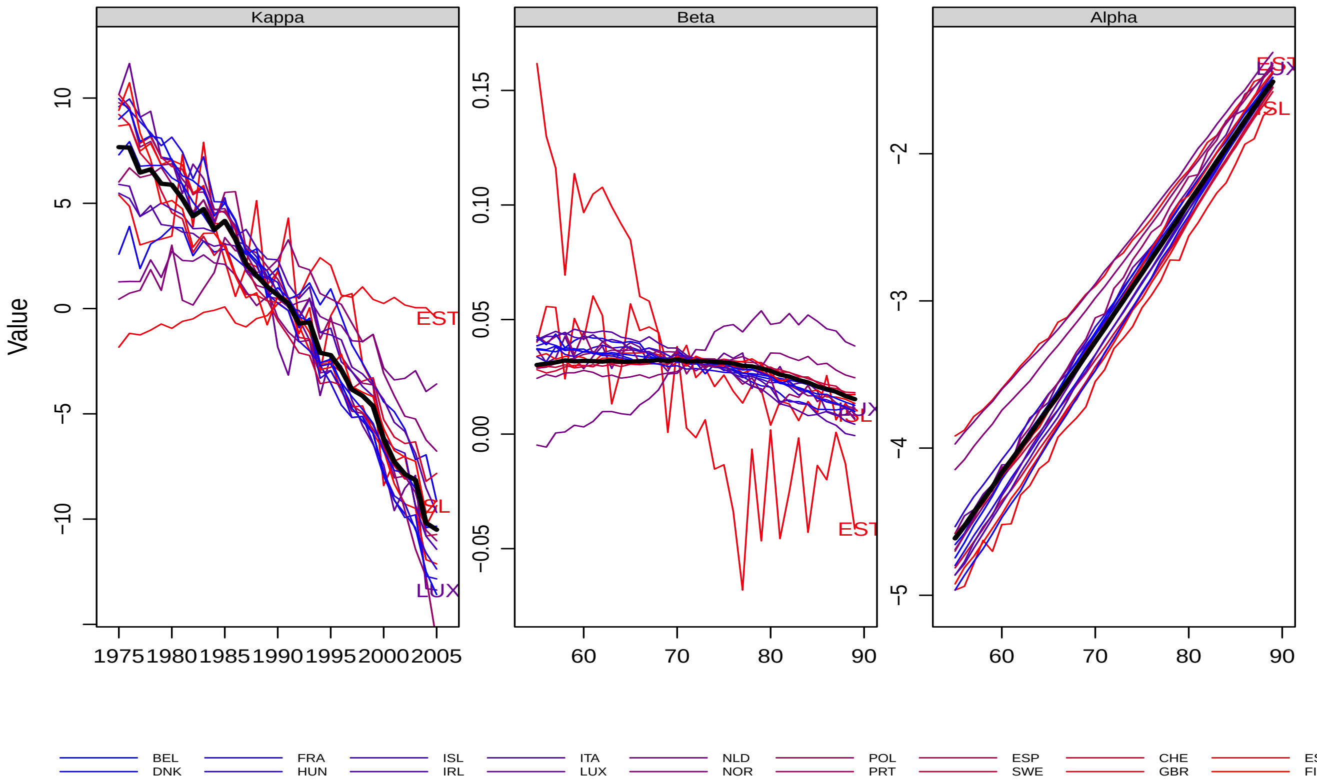

Figure 1: *The estimated parameters* α_x^i, β_x^i *and* κ_t^i *for the 18 countries as well as the aggregate parameters* $\alpha_x^\bullet, \beta_x^\bullet$ *and* $\kappa_t^\bullet$ *in black.*

the procedure described in Russolillo et al. (2011) based on the so-called *Tucker 3*. The Tucker 3 method is also known as the natural extension of principal component analysis for three-way data, i.e. age, time and population, see Russolillo et al. (2011) and Giordano et al. (2019) for more details. In Figure 1, we depicted the corresponding estimations as well as the population specific parameters in the single population model in Equation (2.1), i.e. α_x^i, β_x^i and κ_t^i. We recall that the selected countries corresponds to those with available data over the desired period. However, as we can see in Figure 1, these countries have experienced different mortality improvements over the period. If we look at the evolution of the time-dependent factor κ^i in Figure 1, we see that some countries are exhibit a different pattern. In fact, as we can see in the first column of Table 1, the average decrease of mortality approximated as the average of the decrement of κ^i, i.e. $\mathbb{E}[\Delta\kappa^i]$, is positive. For instance, Estonia and Hungary have known an overall increase of mortality over the consider period. However, as we can in Figure 1, the corresponding β_x^i for these countries is very erratic and takes negative values. This is mainly due, as discussed in Section 3 to the size of these populations. In fact, these two countries represent respectively 0.38% $(266, 258$ exposure-to-risk) and 3.01% $(2, 101, 963$ exposure-to-risk) of the aggregate population.

4.3 Mortality's Forecasts and Credibility Adjustment. In the following, we will not focus

Table 1: *Population specific characteristics over the estimation period* 1975 − 2005 *and ages* 55 − 95. *We report the average decrease of mortality* $\mathbb{E}[\Delta\kappa_t^i]$, *the exposure-to-risk and the number of deaths over the period,* $E^i_{\bullet,\bullet}$ *and* $D^i_{\bullet,\bullet}$, *as well as the relative size compared to the aggregate population.*

	$\mathbb{E}[\Delta\kappa_t^i]$	Exposures		Deaths	
		$E^i_{\bullet,\bullet}$	$E^i_{\bullet,\bullet}/E^{\bullet}_{\bullet,\bullet}(\%)$	$D^i_{\bullet,\bullet}$	$D^i_{\bullet,\bullet}/D^{\bullet}_{\bullet,\bullet}(\%)$
BEL	-0.90385	2,220,401	3.18%	48,335	3.30%
CHE	-0.97535	1,533,425	2.20%	26,327	1.80%
DNK	-0.55322	1,157,804	1.66%	25,533	1.74%
ESP	-0.73799	8,381,679	12.00%	154,255	10.54%
EST	0.05750	266,258	0.38%	7,669	0.52%
FIN	-1.01264	1,070,255	1.53%	21,654	1.48%
FRA	-0.94483	11,924,920	17.07%	238,996	16.32%
GBR	-1.03243	12,434,604	17.80%	272,356	18.60%
HUN	0.18475	2,101,965	3.01%	65,393	4.47%
ISL	-1.00579	49,190	0.07%	771	0.05%
ITA	-1.16494	12,804,063	18.33%	250,967	17.14%
LUX	-1.20779	89,894	0.13%	1,773	0.12%
NLD	-0.79229	3,202,909	4.59%	58,603	4.00%
NOR	-0.86319	932,387	1.33%	19,085	1.30%
POL	-0.29149	6,929,435	9.92%	171,078	11.69%
PRT	-0.83511	2,058,053	2.95%	45,517	3.11%
SWE	-0.93168	2,000,512	2.86%	41,183	2.81%

on the goodness-of-fit performance as the proposed methodology aims at adjusting future forecasts of mortality. Also, the models considered so far are generally over-fitting the data especially when these are of limited amount. We should also recall that the credibility adjustment is, generally, known to produce smoother paths and thus tends to be less favorable when considering mean squared errors as a goodness-of-fit indicator. It is even less adequate knowing that the models used hereafter are based on a minimization of the squared errors and do fit better the realization but, as already mentioned, tend to overfit that data. Also, the use of the explained variation, for instance, is not either a satisfactory diagnostic indicator. Therefore, we will only focus on the predictive performance of the proposed methodology. Hence, in order to forecast mortality we need to model the time-varying parameters κ_t^i and $\kappa_t^\bullet$ for the considered models, see Equations (2.1) and (2.3). Here, we model these time-series using a random walk with drift, although it is not always the optimal model. Hence, our purpose is to adjust the fit in order to coherently forecast the mortality and take into account the credibility of the information stemming from each population. To this end, we should implement the credibility methodology described in Subsection 2.2. Then, using Equation (3.9), the forces of mortality can be successively updated. However, in order to adequately implement the adjustment, we will need to specify the weights ω_t^i. In fact, due to

Assumption **(H5)**, the conditional variance $\mathbb{E}[X_t^i|\Theta_i]$ is assumed to correspond weights as the reciprocals of weights. Obviously, the weights ω_t^i should translate that the information stemming from each population increases with the sample size. Thus, each of the past realization for all populations has a different degree of contribution to the value of the overall credibility estimate, see Equation (3.6). Accordingly, weights ω_t^i of each population will be chosen to reflect the sample size and thus be defined as the expected deaths, by the common factor model in Equation (2.4), for each $t = t_0, \cdots, T$, as follows:

$$\omega_t^i = \sum_{x=\underline{x}}^{\overline{x}} E_{x,t}^i \exp\left(\widehat{\alpha}_x^i + \widehat{\beta}_x^{\bullet}\widehat{\kappa}_t^{\bullet}\right).$$

By doing so, credibility estimators in Equation (3.9), will allocate more weight to population with large sample sizes and thus give the historical particularities of each country due to their weight in projecting population-specific trends. In the last column of Table 2, we report the the factor Z_i. We can see that the latter does convey adequately the sample size and thus the credibility of underlying realization, see also Table 1. We can notic e that the credibility factors Z_i are different depending on the populations considered (from 1.56% for Iceland to 85.42% for United Kingdom). Small portfolios (e.g Ireland, Iceland, Luxembourg, etc.) have small credibility factors whereas big portfolios (France, Italy, etc.) have high credibility factors. This is consistent with what we expected as the credibility factor reflects the individual weight of each population.

4.4 Forecasts and Predictive Performance. The model described above will be used to predict future evolution of mortality. We look at the mortality rates predicted by our adjusted Lee and Carter's model (the joint-κ extension), the Lee and Carter's model in (2.1) and the multi-population model Li and Lee. As noted above, we only validate the performance based on an out-of-sample analysis instead of an in-sample goodness-of-fit inspection. Therefore, in order to validate the behavior and predictive performance of the considered models we fit these to data on the period ranging from 1975 to 2005 for each population, i.e. $t_0 = 1975$ and $T = 2005$, and forecast the future mortality from year 2006 to 2014. In order to measure the forecast error between the true mortality rate and the forecast, we use the mean absolute forecast error (MAFE) and the root squared of the mean forecast error (RSMFE). Let us note h the forecast horizon, then the MAFE is the average of the absolute values of the deviations from the observations $\hat{m}_{x,t}^i$ and it is defined as follows:

$$\text{MAFE}_h^i = \frac{1}{\overline{x} - \underline{x} + 1} \sum_{x=\underline{x}}^{\overline{x}} \frac{1}{h} \sum_{t=T+1}^{T+h} \left|\widehat{m}_{x,t}^i - m_{x,t}^i\right|. \tag{4.1}$$

In the other hand, the RSMFE is defined, for each population $i = 1, \cdots, I$, as follows:

$$\text{RSMFE}_h^i = \sqrt{\frac{1}{\overline{x} - \underline{x} + 1} \sum_{x=\underline{x}}^{\overline{x}} \frac{1}{h} \sum_{t=T+1}^{T+h} \left(\widehat{m}_{x,t}^i - m_{x,t}^i\right)^2}. \tag{4.2}$$

Table 2 reports the two indicators over the projection period, i.e. $h = 9$ years. Also, we reported the estimated credibility factor Z_i in Equation (3.6). We recall that this factor represents

Table 2: *The forecasting performance over the period 2006 − 2014 of the credibility-adjusted model in Equation (2.5) (Adj. LC), the Lee and Carter's model (LC) and the Li and Lee's model (Li-Lee) measured by the MAFE (4.1) and RSMFE (4.2). The grayed cells correspond to the best performing model for each population. The last column reports the credibility factor Z_i over the estimation period.*

	MAFE			RSMFE			Cred. Factor
	Adj. LC	**LC**	**Li-Lee**	**Adj. LC**	**LC**	**Li-Lee**	
BEL	0.0039479	0.0066063	0.0044472	0.0075421	0.0114117	0.0079154	50.08%
CHE	0.0030867	0.0034824	0.0034473	0.0053204	0.0065974	0.0062716	35.73%
DNK	0.0044733	0.0069685	0.0049906	0.0085870	0.0116780	0.0089470	35.23%
ESP	0.0028921	0.0042206	0.0057681	0.0050731	0.0071039	0.0093680	75.36%
EST	0.0130390	0.0255799	0.0155939	0.0268434	0.0367188	0.0335866	12.00%
FIN	0.0043097	0.0061474	0.0047234	0.0077681	0.0111378	0.0081726	29.19%
FRA	0.0031964	0.0031987	0.0040531	0.0061501	0.0058351	0.0070268	82.84%
GBR	0.0048371	0.0049683	0.0054540	0.0072597	0.0078574	0.0083183	85.42%
HUN	0.0080575	0.0087945	0.0071026	0.0160003	0.0146314	0.0142104	55.41%
ISL	0.0159587	0.0157395	0.0236250	0.0338259	0.0309851	0.0661474	1.56%
ITA	0.0022179	0.0024656	0.0025741	0.0034413	0.0038017	0.0042013	83.97%
LUX	0.0148173	0.0154327	0.0164814	0.0331483	0.0350023	0.0358124	3.45%
NLD	0.0049394	0.0116690	0.0075943	0.0074538	0.0183802	0.0113452	54.84%
NOR	0.0048244	0.0075202	0.0056167	0.0100730	0.0136227	0.0103335	29.50%
POL	0.0053159	0.0055251	0.0076762	0.0086950	0.0089433	0.0134398	75.35%
PRT	0.0039544	0.0045444	0.0052898	0.0069503	0.0075475	0.0092128	47.37%
SWE	0.0028745	0.0041731	0.0024075	0.0058471	0.0076027	0.0050234	47.68%

an individual weight on historical observations stemming from population i and takes values in $[0, 1]$. For each portfolio i the larger the volume of historical data, the larger Z_i will be (close to 1). As we can see in the last panel of Table 2, the estimated credibility factor is in line with the size of the population compared to its neighbors, see Table 1. In fact, for small populations this factor assign a small credibility the information flow stemming from the sole population and will favor to learn from the others and thus increase the potential to improve mortality prediction of individual populations. As we can see, the model with credibility (Adj. LC) improves the forecasts in terms of the MAFE for most of the portfolios except for Hungary, Iceland and Sweden. This model also improves the forecasts in terms of the RSMFE except for four countries out of 17. We may note that Iceland are the is country with the least deaths observed, with sometimes few or no deaths at some ages. It represents 0.07% of the global exposure. This may explain why the model with credibility does not forecast better than the model without credibility for these countries. However, the difference in terms of the MAFE between the adjusted model and the classic Lee and Carter's model is less than 1.4%. For Hungary and Sweden, the credibility factor is approximately 50% but the credibility estimator is the closest to 1, see Figure 2. Thus, there are few differences

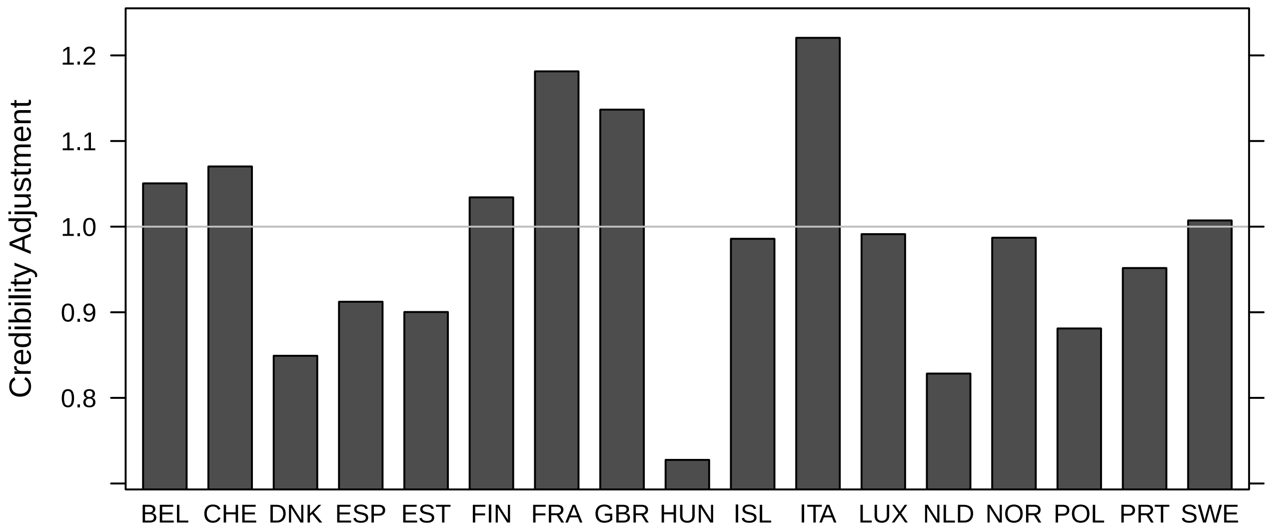

Figure 2: *Next-period adjustment of the joint-κ model based on the credibility adjustment in Equation (3.8)*

between models with and without credibility, which may explain why the model with credibility adjustment does not forecast better either for these countries. For the remaining countries, the adjusted model with credibility estimators (Adj. LC) gives the best results in terms of the MAFE and the RSMFE, compared to the simple Lee and Carter and Li and Lee. In particular, Estonia, Norway, Netherlands, Denmark and Belgium have MAFE and RSMFE that are almost half for a model with credibility compared to the Lee and Carter's model, and therefore benefit very strongly from the learning effect. It is interesting to observe that these countries have low credibility factors but not lower than 11%, see the right panel of Table 2. As soon as the Li and Lee is concerned, we can see that the adjusted model enhance the prediction up to 60% for the Finland, for instance. For the other population the same remark holds. Finally, when the RSMFE is concerned the same conclusion can be drawn, except for France where the adjusted model produce forecast very close to the Lee and Carter's model. In order to understand the performance of the adjusted model for the three populations mentioned above, i.e. Hungary, Iceland and Sweden, we depict the single age mortality evolution over the validation period $2006 - 2014$. Figure 3 displays the historical observations and the projected medians of the three models. At first sight, we see that the adjusted model provides forecasts very close to the Li and Lee (Li-Lee). For Hungary, there is no tendency to a general out-performance of the adjustment except for low ages, i.e. 65 and 75. However, even if the model is not outperforming, the outputs seem to be more prudent that the benchmarks as it forecast lower mortality rates and thus a steady decrease of mortality over time. As regard the Sweden population, we can also see from Figure 3 that the considered models have similar performance. However, at high ages the adjusted model and the Li-Lee models outperforms the Lee-Carter. In Figure 4, we use the forecasts provided by the out-of-sample analysis and derive the

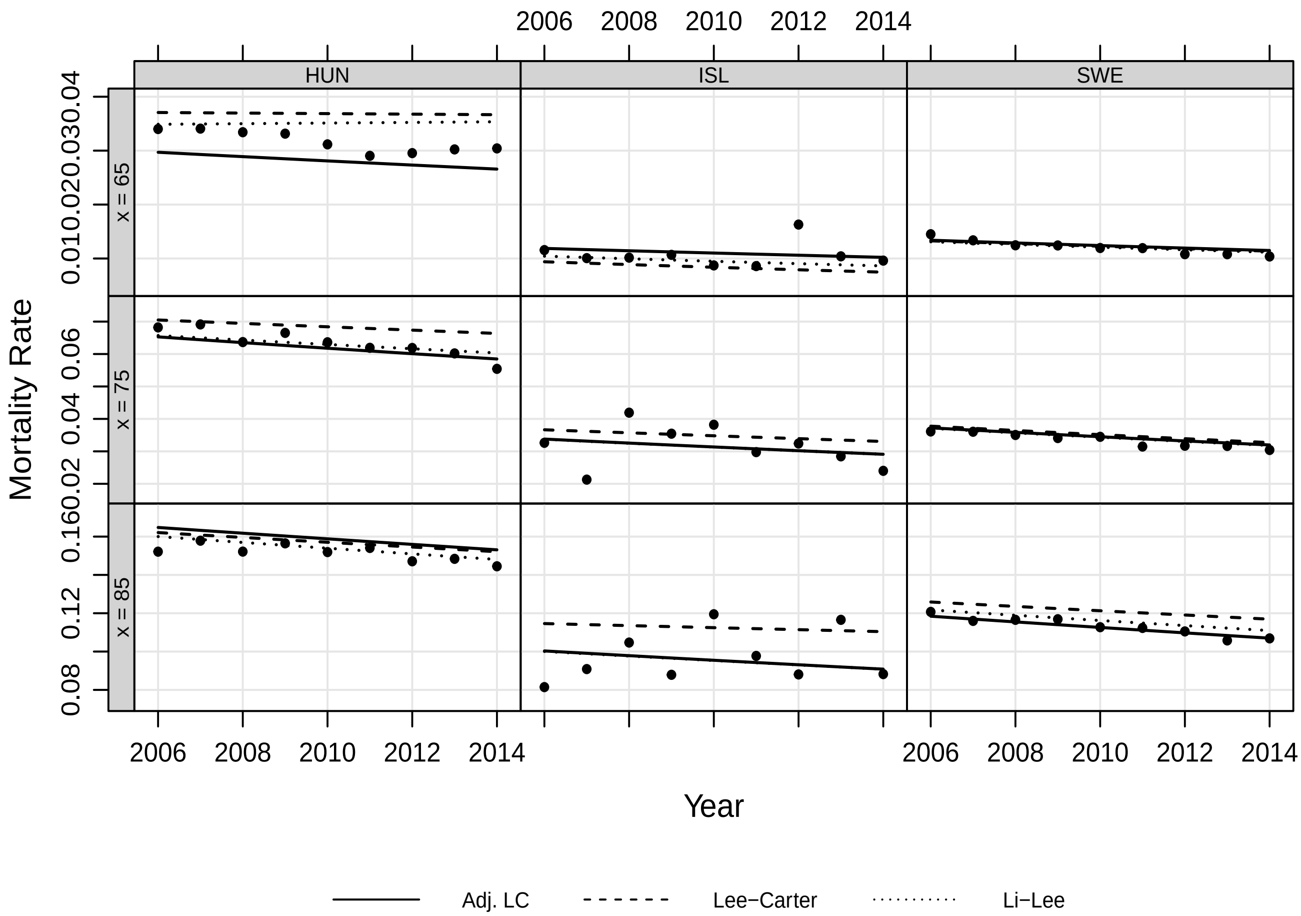

Figure 3: *Mortality rates over the period 2006-2014: Observed mortality (black circles), credibility adjusted forecast (solid line), Lee and Carter's forecast (dashed line) and the Li and Lee (dotted line), for ages* 65, 75 *and* 85.

corresponding projected remaining period life expectancies for ages 65, 75 and 85. We can see that the adjusted model produces forecasts at least equivalent to the benchmarks models. For some ages, the adjustment gives more accurate predictions closer to the observed life expectancies. For these three populations, unlike the forecasting performance of the crude mortality the credibility adjustment model outperforms clearly the benchmarks when it comes to life expectancy prediction.

5 Conclusion

Adequately forecasting mortality is essential in a context of aging societies where people are living increasingly older, in particular for those bound to pay lifetime annuities, such as insurers, pension funds and investors. The increase of prudential regulations, such as Solvency II in Europe, also advocates for a better fit of mortality models in order to better control the risks inherent in

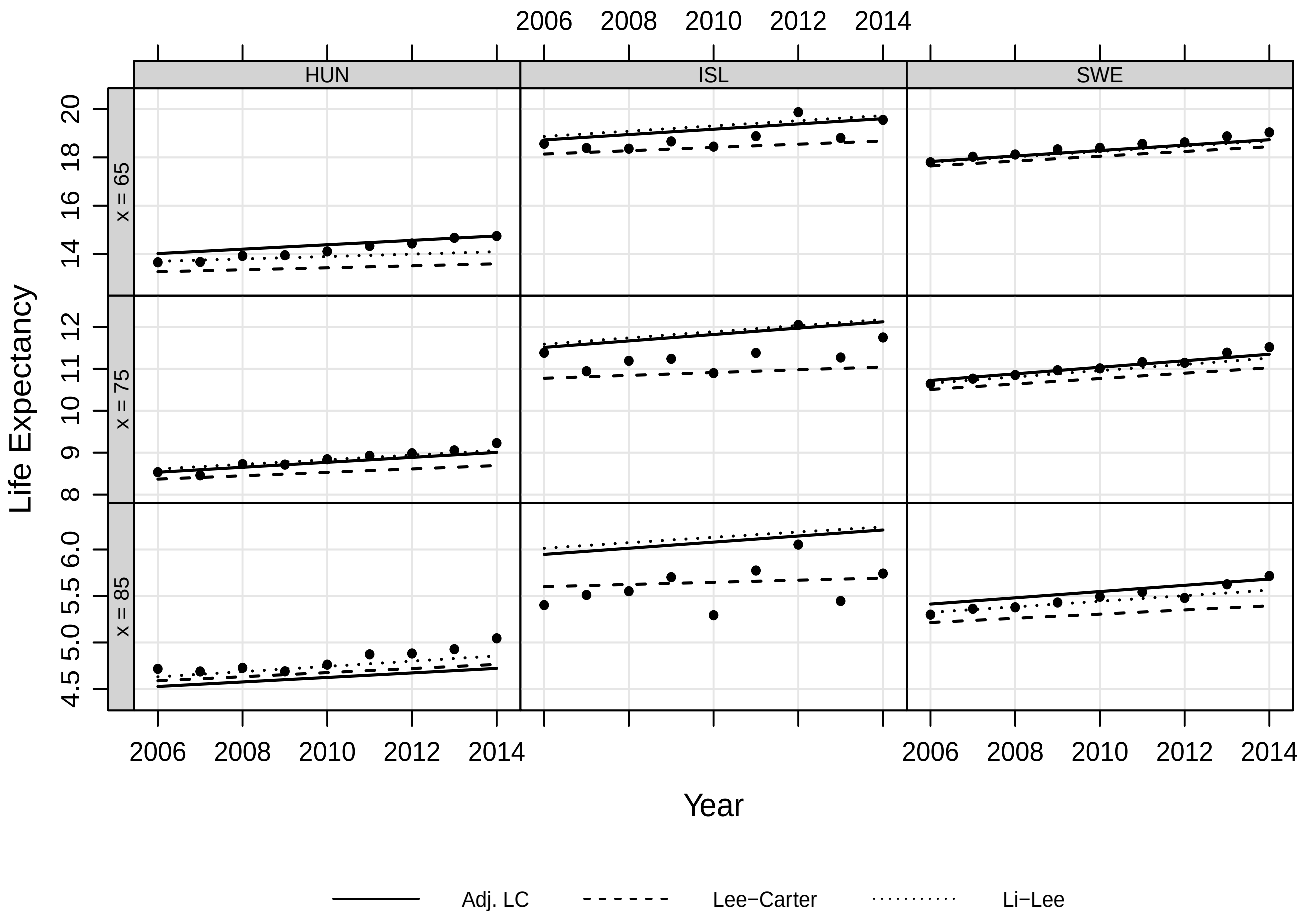

Figure 4: *Life expectancy over the period 2006-2014: Observed life expectancy (black circles), credibility adjusted forecast (solid line), Lee and Carter's forecast (dashed line) and the Li and Lee (dotted line), for ages* 65, 75 *and* 85.

mortality and longevity. In the case of small populations, specific issues arise, and thus more precise models need to be developed.

In this context, this article analyzes the augmented models based on the Lee and Carter's approach in order to better estimate and quantify the longevity risk while relying on a simple interpretation of the parameters dissociating the level risk from the trend risk. It also proposes a specific model that allows to better predict the mortality in the case of small populations. Based on a multiple populations framework, the so-called joint-κ model, the proposed methodology takes into account the data of each small population while relying on an aggregate model giving coherent forecasts. The model, that is the credibility-adjusted, relies on the existing literature but differs from generic models such as relational ones as it takes considers the structure and the data specific to each population while allowing for a learning mechanism from the other populations. Besides, the weight of each population in the aggregate structure is taken into account thanks to the

Bühlmann-Straub credibility framework.

Concretely, a Bühlmann-Straub credibility estimator is applied to the trend of the Lee and Carter's model, allowing to improve future forecasts that will be successively updated and offering more modelling possibilities than a simple Bühlmann credibility estimator. Based on a set of populations data coming from the Human Mortality Database, the results of the paper show that for most countries, the credibility-adjusted Lee and Carter's model gives better forecasts in terms of mortality and life expectancy than a classical Lee and Carter's model or a Li and Lee's model.

This credibility-adjusted model gives results that are easy to interpret thanks to the Lee and Carter's framework and can be applied to lot of different studies involving sub-populations with particular characteristics. In fact, it is possible to forecast deaths of small countries or regions, mortality by cause of death, insurance portfolios specific mortality or sectorial pension funds mortality. Such studies are full of promise, and the model developed in this article is an additional tool to carry them out.

Acknowledgments

Y. Salhi's work is supported by the BNP Paribas Cardif Chair DAMI "Data Analytics & Models for Insurance". The views expressed herein are the author's owns and do not reflect those endorsed by BNP Paribas.

References

Barrieu, P., Bensusan, H., El Karoui, N., Hillairet, C., Loisel, S., Ravanelli, C., and Salhi, Y. (2012). Understanding, modelling and managing longevity risk: key issues and main challenges. *Scandinavian actuarial journal*, 2012(3):203–231.

Bühlmann, H. and Straub, E. (1970). Glaubwürdigkeit für schadensätze. *Bulletin of the Swiss Association of Actuaries*, 70(1):111–133.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., and Khalaf-Allah, M. (2011a). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, 41(1):29–59.

Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2011b). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, 48(3):355–367.

Carter, L. R. and Lee, R. D. (1992). Modeling and forecasting us sex differentials in mortality. *International Journal of Forecasting*, 8(3):393–411.

Coughlan, G., Khalaf-Allah, M., Ye, Y., Kumar, S., Cairns, A., Blake, D., and Dowd, K. (2011). Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. *North American Actuarial Journal*, 15(2):150–176.

Cox, D. (1992). Regression models and life-tables. In *Breakthroughs in statistics*, pages 527–541. Springer.

Danesi, I. L., Haberman, S., and Millossovich, P. (2015). Forecasting mortality in subpopulations using Lee–Carter type models: A comparison. *Insurance: Mathematics and Economics*, 62:151–161.

Debón, A., Montes, F., and Martínez-Ruiz, F. (2011). Statistical methods to compare mortality for a group with non-divergent populations: an application to spanish regions. *European Actuarial Journal*, 1(2):291–308.

Delwarde, A., Kachakhidze, D., Olié, L., and Denuit, M. (2004). Modèles linéaires et additifs généralisés, maximum de vraisemblance local et méthodes relationnelles en assurance sur la vie. *Bulletin Francais d'Actuariat*, 6:77–102.

Doukhan, P., Pommeret, D., Rynkiewicz, J., and Salhi, Y. (2017). A class of random field memory models for mortality forecasting. *Insurance: Mathematics and Economics*, 77:97–110.

Doukhan, P., Rynkiewicz, J., and Salhi, Y. (2020). Optimal neighborhoods selection for AR-ARCH random fields with application to mortality.

Enchev, V., Kleinow, T., and Cairns, A. J. (2017). Multi-population mortality models: fitting, forecasting and comparisons. *Scandinavian Actuarial Journal*, 2017(4):319–342.

Giordano, G., Haberman, S., and Russolillo, M. (2019). Coherent modeling of mortality patterns for age-specific subgroups. *Decisions in Economics and Finance*, 42(1):189–204.

Hainaut, D. (2018). A neural-network analyzer for mortality forecast. *ASTIN Bulletin: The Journal of the IAA*, 48(2):481–508.

Hardy, M. R. and Panjer, H. H. (1998). A credibility approach to mortality risk. *ASTIN Bulletin: The Journal of the IAA*, 28(2):269–283.

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, 63:147–152.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659–671.

Li, H. and Lu, Y. (2018). A Bayesian non-parametric model for small population mortality. *Scandinavian Actuarial Journal*, 2018(7):605–628.

Li, J. S.-H. and Hardy, M. R. (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal*, 15(2):177–200.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography*, 42(3):575–594.

Li, N., Lee, R., and Tuljapurkar, S. (2004). Using the Lee–Carter method to forecast mortality for populations with limited data. *International Statistical Review*, 72(1):19–36.

Renshaw, A. and Haberman, S. (2003). Lee–carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33(2):255–272.

Russolillo, M., Giordano, G., and Haberman, S. (2011). Extending the Lee–Carter model: A three-way decomposition. *Scandinavian Actuarial Journal*, 2011(2):96–117.

Salhi, Y. and Loisel, S. (2017). Basis risk modelling: A cointegration-based approach. *Statistics*, 51(1):205–221.

Salhi, Y. and Thérond, P.-E. (2018). Age-specific adjustment of graduated mortality. *ASTIN Bulletin: The Journal of the IAA*, 48(2):543–569.

Salhi, Y., Thérond, P.-E., and Tomas, J. (2016). A credibility approach of the makeham mortality law. *European Actuarial Journal*, 6(1):61–96.

Schinzinger, E., Denuit, M. M., and Christiansen, M. C. (2016). A multivariate evolutionary credibility model for mortality improvement rates. *Insurance: Mathematics and Economics*, 69:70–81.

Tsai, C. C.-L. and Lin, T. (2017a). A Bühlmann credibility approach to modeling mortality rates. *North American Actuarial Journal*, 21(2):204–227.

Tsai, C. C.-L. and Lin, T. (2017b). Incorporating the Bühlmann credibility into mortality models to improve forecasting performances. *Scandinavian Actuarial Journal*, 2017(5):419–440.

Villegas, A. M. and Haberman, S. (2014). On the modeling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in england. *North American Actuarial Journal*, 18(1):168–193.

Villegas, A. M., Haberman, S., Kaishev, V. K., and Millossovich, P. (2017). A comparative study of two-population models for the assessment of basis risk in longevity hedges. *ASTIN Bulletin: The Journal of the IAA*, 47(3):631–679.

Yue, J. C., Wang, H.-C., and Wang, T.-Y. (2019). Using graduation to modify the estimation of Lee–Carter model for small populations. *North American Actuarial Journal*, pages 1–11.