

Article

Parsimonious Predictive Mortality Modeling by Regularization and Cross-Validation with and without Covid-Type Effect [†]

Karim Barigou ^{*}, Stéphane Loisel  and Yahia Salhi 

ISFA, LSAF EA2429, Univ Lyon, Université Claude Bernard Lyon 1, 50 Avenue Tony Garnier, F-69007 Lyon, France; stephane.loisel@univ-lyon1.fr (S.L.); yahia.salhi@univ-lyon1.fr (Y.S.)

* Correspondence: karim.barigou@univ-lyon1.fr

[†] This work is mainly supported by the AXA research fund project on “Niveaux et tendances d’amélioration de la longévité: sélection de modèle pour les hypothèses “Best Estimate” et détection de rupture à l’aide de tests séquentiels, avec prise en compte du contexte covid-19”. S. Loisel and Y. Salhi also acknowledge support from the BNP Paribas Cardif Chair “Data Analytics & Models for Insurance” and the Milliman research initiative “Actuariat Durable”.

Abstract: Predicting the evolution of mortality rates plays a central role for life insurance and pension funds. Standard single population models typically suffer from two major drawbacks: on the one hand, they use a large number of parameters compared to the sample size and, on the other hand, model choice is still often based on in-sample criterion, such as the Bayes information criterion (BIC), and therefore not on the ability to predict. In this paper, we develop a model based on a decomposition of the mortality surface into a polynomial basis. Then, we show how regularization techniques and cross-validation can be used to obtain a parsimonious and coherent predictive model for mortality forecasting. We analyze how COVID-19-type effects can affect predictions in our approach and in the classical one. In particular, death rates forecasts tend to be more robust compared to models with a cohort effect, and the regularized model outperforms the so-called P-spline model in terms of prediction and stability.

Keywords: mortality; forecasting; regularization; elastic-net; smoothing; Poisson generalized linear model



Citation: Barigou, K.; Loisel, S.; Salhi, Y. Parsimonious Predictive Mortality Modeling by Regularization and Cross-Validation with and without Covid-Type Effect. *Risks* **2021**, *9*, 5. <https://dx.doi.org/10.3390/risks9010005>

Received: 18 November 2020

Accepted: 16 December 2020

Published: 24 December 2020

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

During recent decades, many developed countries have encountered a steady increase in their life expectancy. In this context, the longevity risk, i.e., the risk that people live longer than expected, poses a challenge, not only for public retirement systems planning, but also for the private life annuities business. For Best Estimate of Liabilities and Solvency Capital Requirement computations in Solvency II, the development of statistical tools for the modelling and projection of mortality rates is essential, see [Barrieu et al. \(2012\)](#) for a discussion. A large, international insurer may use different longevity models for different countries and zones. The supervisor sometimes questions model selection. It is natural to wonder whether a data-driven approach can outperform classical models in most countries, or whether it is more reasonable to use distinct models for different zones, in general and in the present COVID-19 context, which breeds unusual mortality rate patterns.

The most influential paper of stochastic modelling is certainly the mortality model proposed by [Lee and Carter \(1992\)](#). This factor-based model decomposes the (logarithmic) age-period surface of mortality patterns into a bilinear combination of age and period parameters. The future period parameters are then forecast by using time series techniques. Several extensions and alternatives were proposed in the literature. For instance, [Renshaw and Haberman \(2006\)](#) extended the Lee–Carter model with the inclusion of a cohort term. The inclusion of a cohort effect improved the fit to the data; however a better fit to data does not necessarily lead to better forecasts of mortality. Indeed, the

projection of the cohort effects can potentially lead to poorer forecasts, see [Hunt and Villegas \(2015\)](#) for a discussion. Other alternatives include the two-factor Cairns–Blake–Dowd (CBD) model introduced by [Cairns et al. \(2006\)](#) and its extensions considered in [Cairns et al. \(2009\)](#), incorporating the quadratic age term and cohort effect. Regime-switching models for mortality modelling were introduced by [Milidonis et al. \(2011\)](#) and later applied in [Hainaut \(2012\)](#) and [Gylys and Šiaulys \(2020\)](#) to accommodate different mortality regimes due to the effect of wars or pandemics. We can also refer to the recent work of [Doukhan et al. \(2017\)](#); [Doukhan et al. \(2020\)](#) and [Guibert et al. \(2019\)](#) for more flexible modeling approaches. In particular, regularization with an elastic-net penalty is also considered in [Guibert et al. \(2019\)](#) to determine the dependence structure between mortality rates within a high-dimensional vector autoregressive approach.

Some papers have considered a quantitative comparison of stochastic mortality models, for instance [Cairns et al. \(2009\)](#) and [Haberman and Renshaw \(2011\)](#). In these studies, the focus is on the ability of the models to fit historical data and not on their ability to predict future mortality. In this paper, we propose a mortality model based on regularization and cross-validation. Regularization is performed by penalizing the Poisson log-linear model based on the elastic-net ([Zou and Hastie \(2005\)](#)), which allows one to prevent overfitting and obtain a sparse model. The elastic-net penalty is a combination of an L_2 -term (as in a ridge regression) to avoid ill-conditioning matrices and an L_1 -term (as in the LASSO regression) to produce a sparse model with a small number of non-zero parameters. Moreover, compared to standard maximum likelihood estimation, the use of cross-validation will automatically select the model which has the best predictive performance. To the best of our knowledge, this paper is the first to apply the elasticnet regularization for smoothing and forecasting the mortality surface. Moreover, when compared to the standard age-period-cohort approach, the advantages of our model are essentially two-fold:

- Standard mortality models have a large number of parameters (easily between 100 and 200 parameters) in comparison with the small number of observations (often around 2000). Regularization techniques allow us to obtain a parsimonious model with very few parameters.
- Model choice is often performed based on the Bayes information criterion (BIC), which measures the ability of the model to fit the historical data (*in-sample criterion*). Our method, based on cross-validation, will select the model which has the best performance in terms of predictions (*out-of-sample criterion*).

The proposed approach uses regularization techniques for smoothing and forecasting mortality rates. To do this, we use a parsimonious decomposition of the mortality surface on a polynomial basis which meant to capture the main features driving the mortality pattern. This can naturally invoke the celebrated spline decomposition that is well-documented in the actuarial and demographic literature. For instance, we can refer to [Currie et al. \(2004\)](#) that proposed a two-dimensional P-spline model for smoothing the mortality surface and forecasting death rates, see also [Camarda et al. \(2012\)](#). While the P-spline and our model both try to smooth and extrapolate the mortality surface, they significantly differ in the choice of the basis and the penalty parameter. Hence, in Section 5.3, we identify robustness issues in the P-spline model and show that our model provides more stable mortality predictions and handles the temporary disruption that can affect the mortality dynamics, e.g. in the context of COVID-19. Regarding the supervisor's search of "one model fits all countries", our results confirm that it is better to use different models for different zones, even if our proposed approach performs reasonably well for all countries and zones.

The rest of the paper is organized as follows. In Section 2, we describe our notation and data at our disposal. Then, in Section 3, we briefly consider the standard approach to mortality modeling based on an age-period-cohort factor decomposition approach. Section 4 explains our approach based on the smoothing of the mortality surface by regularization. Finally, in Section 5, we provide a detailed comparison of our approach to standard mortality models in terms of prediction performance and robustness. Section 6 provides concluding remarks.

2. Notation and Data

We use data on male deaths and exposures from the [Human Mortality Database \(2019\)](#). The data are available by country for (a) single ages from age 0 to 109, with deaths at higher ages gathered into the category 110+. In our analysis, we consider the period 1970–2017 (last year currently available) and the age range 50–90, the range of interest for pension and annuity providers. Moreover, in our comparison of mortality models, we consider the data for USA, UK, France, Belgium, Japan and Hong Kong.

Let the random variable $D_{x,t}$ denote the number of deaths in a population aged x last birthday during calendar year t . Moreover, let $\mathbf{D} = (d_{x,t})$ be the matrix of the observed number of deaths and $\mathbf{E}^c = (e_{x,t})$ be the matrix of observed central exposures, where $e_{x,t}$ is the exposure-to-risk, which measures the average population during calendar year t over the age band $[x, x + 1]$. The force of mortality and central death rates are denoted by $\mu_{x,t}$ and $m_{x,t}$, respectively. We make the standard assumption that the force of mortality is constant over each year of age and calendar year, i.e. $\mu_{x+s,t+u} = \mu_{x,t}$ for all $0 \leq u, s < 1$, so that the force of mortality $\mu_{x,t}$ and death rate $m_{x,t}$ coincide. This assumption has the following two consequences:

1. The mortality rate $q_{x,t}$ which represents the probability that an individual aged exactly x at time t dies in the following year, i.e., between t and $t + 1$, is given by

$$q_{x,t} = 1 - e^{-\mu_{x,t}} = 1 - e^{-m_{x,t}}.$$

2. If we treat the number of deaths $d_{x,t}$ as a random variable $D_{x,t}$ and the central exposure $e_{x,t}$ as fixed, then $D_{x,t}$ follows a Poisson distribution

$$D_{x,t} \sim \text{Poisson}(e_{x,t}\mu_{x,t}), \quad (1)$$

where $\mathbb{E}\left(\frac{D_{x,t}}{e_{x,t}}\right) = \mu_{x,t}$, see [Pitacco et al. \(2009\)](#).

It thus follows that the maximum likelihood estimate $\hat{\mu}_{x,t}$ of the force of mortality is given by

$$\hat{\mu}_{x,t} = \frac{d_{x,t}}{e_{x,t}}.$$

3. Standard Approach to Mortality Modeling

Following [Hunt and Blake \(2014\)](#) and [Currie \(2016\)](#), the majority of mortality models can be expressed as generalized linear or non-linear models. In this section, for the sake of comparison to our regularized model, we briefly discuss the general structure of standard mortality models, their estimations and projections.

3.1. The General Structure of Mortality Models

Most of the existing mortality models fall into an age-period-cohort model. Mathematically, they have the following form:

$$g\left(\mathbb{E}\left(\frac{D_{x,t}}{e_{x,t}}\right)\right) = \alpha_x + \sum_{i=1}^N \beta_x^{(i)} \kappa_t^{(i)} + \gamma_{t-x}. \quad (2)$$

This equation has the following components:

- A link function g transforms the observed data into a suitable linear or non-linear form for modeling. The standard choice under the Poisson assumption (1) is the log link function (see e.g., [Currie \(2016\)](#) for a good justification in the context of mortality modelling).
- The static age function term α_x which captures the average shape of the mortality curve.
- N period-age terms $\beta_x^{(i)} \kappa_t^{(i)}$ ($i = 1, \dots, N$) which describe the mortality trends.
- The term γ_{t-x} represents the lifelong cohort effect of certain generations.

In the remainder of this paper, we always assume a logarithmic link function and a Poisson distribution of deaths. Therefore, Equation (2) reduces to

$$\log \mu_{x,t} = \alpha_x + \sum_{i=1}^N \beta_x^{(i)} \kappa_t^{(i)} + \gamma_{t-x}, \quad (3)$$

where we denote the set of parameters in the model by

$$\phi := (\alpha_x, \beta_x^{(1)}, \dots, \beta_x^{(N)}, \kappa_t^{(1)}, \dots, \kappa_t^{(N)}, \gamma_{t-x}).$$

For instance, in the case of the simplified version of the [Renshaw and Haberman \(2006\)](#) model (Lee–Carter model with a cohort effect), the equation is given by

$$\log \mu_{x,t} = \alpha_x + \beta_x \kappa_t + \gamma_{t-x}.$$

This simplified structure has been suggested by [Haberman and Renshaw \(2011\)](#) to resolve some stability issues of the original model.

3.2. Estimation and Standard Model Choice Criterion

Given the matrix of observed deaths $\mathbf{D} = (d_{x,t})$ and the matrix of observed central exposures $\mathbf{E}^c = (e_{x,t})$, the set of parameters in (3) are estimated by maximizing the log-likelihood, which is given by

$$\mathcal{L}(\hat{\phi}) = \sum_x \sum_t d_{x,t} \log(e_{x,t} \mu_{x,t}) - e_{x,t} \mu_{x,t} - \log(d_{x,t}!), \quad (4)$$

in the case of a Poisson distribution of deaths, see (1), and the intensity $\mu_{x,t}$ given in (3). Since the underlying intensity is non-linear, this maximization is typically performed by using an iterative Newton–Raphson procedure. We remark that several mortality models are non-identifiable, namely several sets of parameters maximize the log-likelihood, and therefore, it is common to introduce additional constraints to ensure the uniqueness of the solution, see for instance [Millossovich et al. \(2018\)](#).

In order to compare the goodness-of-fit of different mortality models, a standard choice consists of computing the Bayes information criterion (BIC) for each model and selecting the model that provides the lowest value¹. The BIC modifies the likelihood by including a penalty on the number of parameters:

$$BIC = \nu \log K - 2\mathcal{L}(\hat{\phi}),$$

where $\mathcal{L}(\hat{\phi})$ is the maximum likelihood estimate, K the number of observations and ν the effective number of estimated parameters. The BIC tries to find a balance between quality of fit (which can always be improved by increasing the number of parameters) and parsimony. Based on the BIC, mortality models with the addition of cohort parameters often lead to the best fitting models, but they however do not necessarily lead to the best forecasts. The choice of identifiability constraints or fitting strategies keeps the quality of fit constant, but may severely impact the projected mortality rates (see e.g., Section 8 in [Currie \(2016\)](#)).

3.3. Forecasting

Projecting mortality rates follows from projecting the period indexes $\kappa_t^{(i)}$ ($i = 1, \dots, N$) and the cohort index γ_{t-x} in Equation (3), by using time series techniques. For the pe-

¹ This type of analysis was for instance performed in Section 6.1.1. in [Cairns et al. \(2006\)](#) or Section 3.3 in [Haberman and Renshaw \(2011\)](#).

riod indexes, it is often assumed that they follow a multivariate random walk with drift (Haberman and Renshaw (2011)):

$$\kappa_t = \delta + \kappa_{t-1} + \zeta_t^k, \quad \kappa_t = \begin{pmatrix} \kappa_t^{(1)} \\ \vdots \\ \kappa_t^{(N)} \end{pmatrix}, \quad \zeta_t^k \sim \mathcal{N}(\mathbf{0}, \Sigma),$$

where δ is a vector of drift parameters and Σ is the variance-covariance matrix of the white noise ζ_t^k . For the cohort index, previous studies assumed that the dynamics of the cohort effect are independent of the dynamics of the period effects $\kappa_t^{(i)}$ (Cairns et al. (2011) and Lovász (2011)). The standard procedure is to consider a full range of ARIMA(p, d, q) for different values of p, d, q and select the optimal values with the BIC. As pointed out by Currie (2016), period and cohort effects are typically not independent; therefore, the independence assumption should be treated with care.

4. Regularized Mortality Modeling

This section outlines how the elastic-net regularization and cross-validation can be used to the smoothing and forecasting of two-dimensional mortality tables.

4.1. Polynomial Representation of Mortality Models

Similar to the previous section, we assume a Poisson model for death counts and a log link function. However, instead of considering the specific structure (3) for the logarithm of the death rates, we investigate the following structure:

$$\log \mu_{x,t} = f(x, t), \quad (5)$$

where $f(x, t) : \mathbb{R}^2 \rightarrow \mathbb{R}$ represents the mortality surface.

We point out that if the function f is continuous on $[a, b]^2$, by the Stone–Weierstrass theorem, for every $\epsilon > 0$, there exists a polynomial p such that $|f(x, t) - p(x, t)| < \epsilon$ for all $x, t \in [a, b]^2$. This result would involve infinitely many parameters, which leaves a direct estimate infeasible. Instead, we truncate the polynomial basis and only consider polynomials up to order n :

$$\log \mu_{x,t} = \sum_i \sum_j a_{ij} x^i t^j, \quad 0 \leq i + j \leq n, \quad (6)$$

where we denote by $\mathbf{a} = (a_{ij})$ the set of parameters to be estimated.²

Evidently, if one performs a standard maximum likelihood estimation by Equation (4) for a high order n , the model will typically be overparameterized and provide poor forecasts. For this reason, we combine techniques of regularization and cross-validation to obtain a parsimonious model which is optimal in terms of predictions.

4.2. Estimation by Regularization and Cross-Validation

In order to reduce the model complexity in Equation (6), we use regularization techniques which penalize the size of the coefficients. Ridge penalizes the L_2 -norm of the coefficients and tends to shrink highly-correlated variables toward one another. Lasso penalizes the L_1 -norm and tends to pick one of them and discard the others, producing a

² Moreover, we consider the orthogonal polynomial basis through the `polym` R function to ensure numerical stability.

sparse model. The elastic-net penalty mixes these two penalties and aims to optimize the following penalized log-likelihood:

$$\max_a \mathcal{L}(a) - \lambda \underbrace{\left((1 - \alpha) \underbrace{\sum_{i,j} (a_{ij})^2}_{\text{Ridge}} + \alpha \underbrace{\sum_{i,j} |a_{ij}|}_{\text{Lasso}} \right)}_{\text{Elastic-net}}, \tag{7}$$

where $\mathcal{L}(a)$ is the log-likelihood of the model, as given in Equation (4), $\alpha \in [0, 1]$ determines the mix between Ridge and Lasso penalties and λ is a positive parameter which measures the strength of the penalties. As λ increases, the number of parameters which are selected decreases and the model becomes more parsimonious.

For the selection of the optimal parameters α and λ , we use a K -fold cross-validation (with $K = 10$): the data are split into K parts; $K - 1$ parts are used to fit the model (called the training set) and we compute the prediction error of the fitted model on the last part (called the validation set). Standard choices to measure the prediction error are the mean-squared-error (MSE) or the deviance. Since we consider a penalized Poisson model, we evaluate the prediction performance via the Poisson deviance (on the validation set), which is given by

$$D(\hat{a}) = 2(\mathcal{L}(\bar{a}) - \mathcal{L}(\hat{a})),$$

where $\mathcal{L}(\bar{a})$ is the log-likelihood of the fitted saturated Poisson model (a model with a free parameter per observation) and $\mathcal{L}(\hat{a})$ is the log-likelihood of the fitted Poisson model.

Once the parameters α and λ which minimize the Poisson deviance on the validation set have been selected, the resulting set of parameters \hat{a} leads to a smoothed mortality surface and provides the best model in terms of out-of-sample prediction performance.

4.3. Forecasting

We treat the forecasting of future mortality rates by extrapolation of the fitted mortality surface. After the estimation procedure described in Section 4.2 by regularization and cross-validation, we obtain the following fitted model:

$$\log \hat{\mu}_{x,t} = \sum_i \sum_j \hat{a}_{ij} x^i t^j, \quad 0 \leq i + j \leq n,$$

where $\hat{a} = (\hat{a}_{ij})$ is the set of the fitted parameters.

We note that by the presence of the L_1 penalty term, most of these coefficients will be equal to zero, which leads to a parsimonious fitted mortality model.

If the model is fitted for the years t_1, \dots, t_{n_1} , and we wish to forecast for n_2 years into the future, we consider the following extrapolation of central projections:

$$\log \hat{\mu}_{x,t} = \sum_i \sum_j \hat{a}_{ij} x^i t^j, \quad 0 \leq i + j \leq n,$$

for $t = t_{n_1+1}, \dots, t_{n_1+n_2}$.

4.4. Effect of COVID-19

In the context of COVID-19, it is important to determine how mortality models and forecasts react to a certain pandemic shock. In the following, we consider two scenarios: one optimistic and one pessimistic, and analyze how our regularized model and the P-spline model change their predictions due to these scenarios.

As a first scenario, we follow the ideas of Cairns et al. (2020). The authors proposed an accelerated deaths model, following the assumption that “Many of those who die from coronavirus would have died anyway in the relatively near future due to these co-

morbidities". Their model depends essentially on two parameters: $\alpha(x)$, which represents the amplitude of COVID-19 deaths at age x and $\rho(x)$, which represents the expected remaining years of life lost by those who die from Covid-19. Hereafter, we propose to perturb deaths in a way that reflects the accelerated deaths model:

- During the year of the pandemic, $t = t_p$, we assume that there is a death increase of $\beta(x)$:

$$d_{x,t}^{\text{new}} = (1 + \beta(x))d_{x,t}.$$

- The next K years of the pandemic, $t = t_p + 1, \dots, t_p + K$, deaths and exposures are reduced to compensate the deaths of COVID-19:

$$\begin{aligned} d_{x,t}^{\text{new}} &= d_{x,t} - \frac{d_{x-m,t_p}}{K}, \\ e_{x,t}^{\text{new}} &= e_{x,t} - d_{x-m,t_p} \frac{K - m + 1}{K}, \end{aligned} \quad (8)$$

where $m := t - t_p = 1, \dots, K$ is the number of years after the pandemic. Equations (8) reflect the idea that people who died from COVID-19 would have died in the next K years anyway. We note that we spread the deaths equally over the K years. To draw a parallel with the accelerated death models, $\beta(x)$ measures the amplitude of the pandemic at age x and K measures the reach of the pandemic.

As a second scenario, we consider a more pessimistic perspective and assume that there is a persistent increase of death rates after the pandemic. This is modelled by

$$d_{x,t}^{\text{new}} = (1 + \beta(x))d_{x,t}, \quad \text{for } t = t_p, t_p + 1, \dots \quad (9)$$

Hence, under this scenario, deaths during and after the pandemic are increased by $\beta(x)$ and do not revert to normal.

4.5. Model Uncertainty and Prediction Intervals

Statistical testing and the obtention of prediction intervals are not directly possible using the elastic net regularization, as no standard errors for the estimated parameters are computed directly. There is some discussion concerning hypothesis testing in the Lasso model (see e.g. [Lockhart et al. \(2014\)](#) and [Tibshirani et al. \(2018\)](#)), but following [Kyung et al. \(2010\)](#), it seems that there is currently no general agreement on a statistically valid method of obtaining standard errors of lasso and elastic net estimates.

When analyzing uncertainty in mortality projections with age-period-cohort models, there are essentially two random sources: on the one hand, the uncertainty in the projection of period and cohort indexes, and on the other hand, the uncertainty from the estimation of the parameters in the model. Typically, the first type of uncertainty is treated with standard time series techniques, while the second type can be taken into account with bootstrap techniques ([Brouhns et al. \(2005\)](#); [Koissi et al. \(2006\)](#); [Renshaw and Haberman \(2008\)](#)).

By construction, our regularized mortality model described in Section 4 only includes the second type of uncertainty, since there are no period and cohort indexes to be projected for the extrapolated mortality rates. We did perform a bootstrap analysis of our model but the obtained prediction intervals were too narrow compared to the ones of standard mortality models. We suspect that the reason for this is that the main uncertainty in standard mortality models comes from the projection of period and cohort indexes, which is not present in our model. In order to obtain prediction intervals which are comparable to standard age-period-cohort models, we introduce hereafter a simple procedure to obtain prediction intervals by the inclusion of a period index.

Let us assume that from a matrix of deaths $\mathbf{D} = (d_{x,t})$ and central exposures $\mathbf{E}^c = (e_{x,t})$, for $x = x_1, \dots, x_k$ and $t = t_1, \dots, t_{n_1}$, we have obtained with the regularized elastic net model fitted and forecast central death rates:

$$\hat{\mu}_{x,t} = \exp\left(\sum_{i,j} \hat{a}_{ij} x^i t^j\right), \quad 0 \leq i + j \leq n, \quad (10)$$

for $t = t_1, t_2, \dots, t_{n_1+n_2}$, where n_1 is the number of years in the data and n_2 the number of forecast years. Our goal is to obtain prediction intervals for forecast death rates which are centered around the extrapolated central death rates obtained in (10). The general procedure is as follows:

1. In order to extract a period index, we assume that mortality rates improve at a general improvement rate λ_t per year:

$$\mu_{x,t+1} = \lambda_t \mu_{x,t}, \quad \text{for any } x = x_1, \dots, x_k.$$

2. We estimate past improvement rates with the simple estimator:

$$\hat{\lambda}_t = \frac{\sum_x \hat{\mu}_{x,t+1}}{\sum_x \hat{\mu}_{x,t}}$$

where $\hat{\mu}_{x,t+1}$ and $\hat{\mu}_{x,t}$ are the smoothed death rates.³

3. We fit a random walk with drift (RWD) to past improvement rates: $\hat{\lambda}_t$ for $t = t_1, \dots, t_{n_1}$.
4. After the RWD fit, we can then obtain prediction intervals for future death rates:

$$IC_{1-\alpha}(\mu_{x,t+1}) = [\lambda_{t,\alpha/2} \hat{\mu}_{x,t}; \lambda_{t,1-\alpha/2} \hat{\mu}_{x,t}]$$

where $\lambda_{t,\alpha/2}$ and $\lambda_{t,1-\alpha/2}$ are the standard quantiles for the improvement rate λ_t for $t = t_{n_1+1}, \dots, t_{n_1+n_2}$.

5. Prediction intervals centered around the extrapolated central death rates can then be obtained with a shift of the mean:

$$IC_{1-\alpha}^c(\mu_{x,t+1}) = [\lambda_{t,\alpha/2} \hat{\mu}_{x,t} - \bar{\lambda}_t \hat{\mu}_{x,t} + \hat{\mu}_{x,t+1}; \lambda_{t,1-\alpha/2} \hat{\mu}_{x,t} - \bar{\lambda}_t \hat{\mu}_{x,t} + \hat{\mu}_{x,t+1}].$$

where $\bar{\lambda}_t$ is the mean of λ_t .

Following this procedure, we obtain prediction intervals that are centered around the central death rates estimated from the regularized elastic net model. The uncertainty around the central death rates follows from the uncertainty of the improvement rate λ_t , which presents the same dynamics as the period index κ_t in the Lee–Carter model.

5. Numerical Analysis and Comparison of Models

This section is dedicated to the numerical analysis of our model and its comparison with standard actuarial models. In Section 5.1, we show how the implementation can be easily performed in the R language and applied to the French data. Section 5.2 compares the prediction performance of our model with standard age-period-cohort models. In Section 5.3, we make a detailed comparison with the P-spline model and particularly show that the regularized model is more robust and outperforms the P-spline approach in terms of prediction errors.

³ The estimated improvement rate on crude death rates is very noisy. The use of smoothed death rates allows one to have a more regular pattern for the improvement rate.

5.1. Smoothing and Forecasting Central Death Rates for France

The previous section set out the methodology to estimate and forecast mortality rates by regularization and cross-validation. We now apply our method to the French data set from HMD. Given the matrix of deaths $\mathbf{D} = (d_{x,t})$ and the matrix of exposures $\mathbf{E}^c = (e_{x,t})$, Figure 1 represents historical death rates which are given by

$$\hat{\mu}_{x,t} = \frac{d_{x,t}}{e_{x,t}}.$$

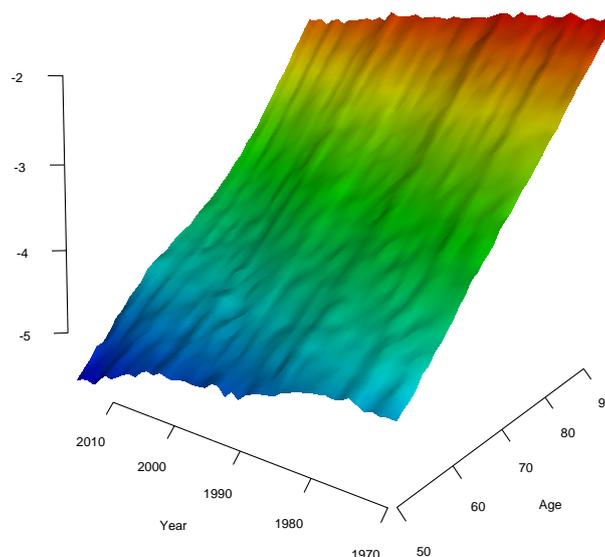


Figure 1. Surface of the logarithm of the raw death rates for age from 50 to 90 and year from 1970 to 2017 for French males.

Our objective is twofold: first, we smooth the mortality surface by regularization and cross-validation. In a second step, we forecast future death rates by extrapolation.

For the smoothing of the mortality surface, we consider the polynomial representation (6) with $n = 20$. Our model is a Poisson generalized linear model with log link function which can be expressed as

$$\log \mathbb{E}(D_{x,t}) = \log e_{x,t} + \sum_{i,j} a_{ij} x^i t^j,$$

where the elasticnet penalization applies on the set of parameters $\mathbf{a} = (a_{ij})$. The estimation by regularization and cross-validation is performed with the use of the R-package **glmnet** (Friedman et al. (2010)).

We performed this cross-validation for a range of α values and kept the model which has the lowest Poisson deviance on the validation set. Figure 2 represents the cross-validation results for the optimal α . The algorithm starts with the constant model (no parameter), and as λ decreases, the number of parameters increases. The algorithm stops when the increase of parameters does not improve the Poisson deviance of the validation set anymore. Figure 2 shows that the algorithm chose a model of 31 parameters (among 230 parameters).

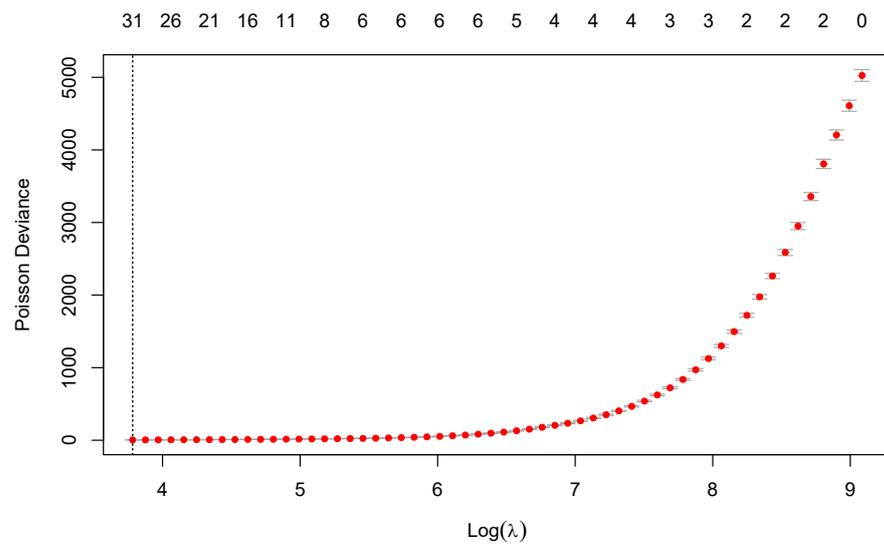


Figure 2. Cross-validation plot of the `glmnet` algorithm for $\alpha = 0.5$.

Forecast of mortality can then be obtained by extrapolation of the fitted mortality surface. Figure 3 represents the crude death rates (the data) and the forecast death rates until 2050. At first glance, we observe that the fitted model with only 31 parameters performs well at forecasting mortality rates. In particular, forecast death rates extend properly the observed levels of mortality rates represented by the different color levels in Figure 3.

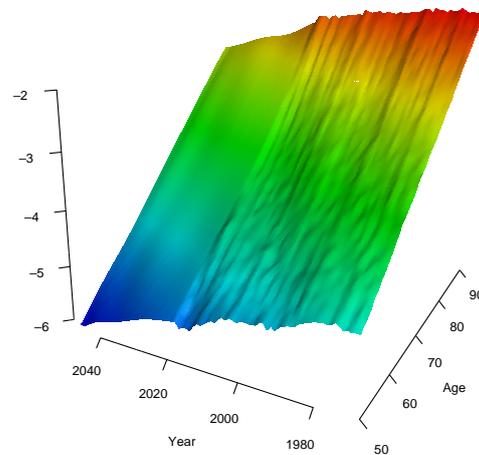


Figure 3. Logarithm of crude death rates from year 1970 to 2017 and logarithm of extrapolated death rates from 2018 to 2050.

5.2. Comparison with Standard Age-Period-Cohort Mortality Models

In this part, we compare our regularized mortality model with a family of popular mortality models in terms of prediction performance. The models considered for comparison are listed in Table 1 and are fitted with the R package `StMoMo` (Millossovich et al. (2018)) in the log-Poisson setting with appropriate identifiability constraints.

Table 1. Model structures considered for comparison.

Mortality Model	Dynamics
Lee–Carter (LC)	$\log \mu_{x,t} = \alpha_x + \beta_x \kappa_t^{(1)}$
Cairns–Blake–Dowd (CBD)	$\log \mu_{x,t} = \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)}$
Age-Period-Cohort (APC)	$\log \mu_{x,t} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x}$
Renshaw–Haberman (RH)	$\log \mu_{x,t} = \alpha_x + \beta_x \kappa_t^{(1)} + \gamma_{t-x}$
M7	$\log \mu_{x,t} = \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)} + ((x - \bar{x})^2 - \bar{\sigma}_x^2) \kappa_t^{(3)} + \gamma_{t-x}$

To measure prediction accuracy, we fitted the models based on data from 1986 to 2007 and compared the difference between the realized death rates and the forecasted death rates from 2008 to 2017 (10-year forecast). To measure this difference, we computed the mean absolute error (MAE), which, in this case, is given by

$$MAE = \sum_{x=50}^{90} \sum_{t=2008}^{2017} \left| \frac{d_{x,t}}{e_{x,t}} - \hat{\mu}_{x,t} \right|$$

where $d_{x,t}$ and $e_{x,t}$ are the realised death and exposures for age x and year t , and $\hat{\mu}_{x,t}$ is the forecasted death rate from the model.

Table 2 reports the MAE for standard mortality models and our regularized model for six countries: the USA, the UK, France, Belgium, Japan and Hong Kong. We highlight for each country the models with the lowest prediction error and the lowest BIC. From Table 2, we observe that the models which have the best fit to historical data (in terms of BIC) may provide poor forecasts (Cairns et al. (2011)). In particular, for France, the inclusion of cohort terms in the Lee–Carter model provided a better fit, but the Lee–Carter model without cohort terms led to better forecasts. Moreover, the regularized model performs among the top 3 in terms of MAE for most countries and prediction results through different countries are more stable for the regularized model compared to mortality models with cohort effects.

Table 2. Mean Absolute Error for USA, UK, France, Belgium (BEL), Japan (JAP) and Hong Kong (HKG) for the different models of Table 1 and our regularized model (Reg). Fit based on historical data from 1986 to 2007. In bolds: Lowest Mean Absolute Error per Country and in italics: lowest BIC model (excluding Reg).

	Mean Absolute Error (10-Year Forecast)					
	USA	UK	France	BEL	JAP	HKG
LC	0.00369	0.00225	0.00120	0.00246	0.00213	0.00265
CBD	0.00279	0.00205	0.00505	0.00388	0.00313	<i>0.00261</i>
APC	0.00161	0.00251	0.00206	0.00338	0.00250	0.00168
RH	<i>0.00368</i>	0.00463	<i>0.00233</i>	0.00308	<i>0.00194</i>	0.00275
M7	0.00222	0.00152	0.00259	0.00200	0.00097	0.00143
Reg	0.00194	0.00212	0.00274	0.00244	0.00267	0.00214

We now assess the robustness of the different models relative to the sample period used in estimating the model. For each country⁴ and each model, we consider historical data from 1970 to 2007 in Table 3 instead of the period 1986 to 2007, as considered in Table 2. Compared to Table 2, we can make the following observations:

- The RH model is very sensitive to the fitted period. The model can go from the worst to the best model (cf. USA) and from the best to the worst model (cf. Belgium). Cairns et al. (2011) already pointed out that lack of robustness in this model as parameter estimates jump to a qualitatively different solution when we use more data.

⁴ Except Hong Kong, as data from Hong Kong are only available from 1986 onwards.

- The regularized model and the Lee-Carter model appear to be reasonably robust relative to changes in the period of data used either in terms of MAE values or ranking positions. In particular, models without cohort effects tend to be more robust towards the sample data used as they do not require the projections of cohort terms.

Table 3. Similar table as Table 2 but based on historical data from 1970 to 2007.

	Mean Absolute Error (10-Year Forecast)				
	USA	UK	France	BEL	JAP
LC	0.00432	0.00318	0.00147	0.00345	0.00213
CBD	0.00262	0.00193	0.00490	0.00335	0.00355
APC	0.00201	0.00265	0.00261	0.00370	0.00328
RH	0.00197	0.00313	0.00293	0.00622	0.00167
M7	0.00407	0.00195	0.00227	0.00142	0.00298
Reg	0.00288	0.00307	0.00262	0.00287	0.00310

5.3. Comparison with the P-Spline Model

Our model relies on the smoothing of mortality over both age and time, where the smoothing is performed by the penalty in the elastic-net regularization (7). In the actuarial literature, Currie et al. (2004) proposed two-dimensional P-splines for smoothing and forecasting mortality rates, and this approach is implemented in the R package **MortalitySmooth** (Camarda et al. (2012)). Hereafter, we highlight similarities and differences with our approach. Moreover, 10-year forecasts of both methods are also compared for different countries and extrapolation issues in the P-spline approach are also discussed. In particular, we found that the P-spline approach is sensitive to small perturbations and provides unrealistic forecasts for the USA and in the context of COVID-19.

The common point between both modeling approaches is the use of a penalized generalized linear model with Poisson errors. However, they have two important differences:

- The basis: In the P-spline model, the basis consists of cubic B-splines, which are bell-shaped curves composed of smoothly joined polynomial pieces of degree 3. By construction, B-splines are defined in a local manner, i.e. each B-spline is non-zero only in a certain neighborhood of the mortality surface. On the other hand, our regularized mortality model uses a global two-dimensional polynomial basis, see (6). More flexibility can be included either by increasing the number of knots in the B-spline approach or by increasing the degree order in the global polynomial basis.
- The penalty: To smooth the mortality surface and avoid overfitting, a penalty term on the regression coefficients is introduced. In the P-spline model, the following penalty matrix is defined:

$$P = P(\lambda_a, \lambda_y) = \lambda_a \mathbf{I}_{c_y} \otimes \mathbf{D}_a^T \mathbf{D}_a + \lambda_y \mathbf{I}_{c_a} \otimes \mathbf{D}_y^T \mathbf{D}_y,$$

where λ_a and λ_y are the smoothing parameters in age and year, respectively, and \mathbf{D}_a and \mathbf{D}_y are second order difference matrices acting on the columns and rows of \mathbf{A} , respectively. Here, \mathbf{A} is the matrix of coefficients arranged in a matrix $c_a \times c_y$, where c_a and c_y are the dimensions of the bases spanning age and year, respectively. Smoothing parameters (λ_a, λ_y) are chosen by minimizing the BIC. More details can be found in Currie et al. (2004).

On the other hand, in our regularized model, the penalty term is given by the elastic-net regularization:

$$P = P(\lambda, \alpha) = \lambda \left((1 - \alpha) \sum_{i,j} (a_{ij})^2 + \alpha \sum_{i,j} |a_{ij}| \right),$$

where λ is the smoothing parameter and α is the mixing parameter between the two types of penalization. An important difference is that both parameters (λ, α) are

chosen in terms of prediction performance by cross-validation on out-of-sample data instead of in-sample criterion such as the BIC.

In Table 4, we compare 10-year forecasts based on historical data from 1986 to 2007 for USA, UK, France, Belgium, Japan and Hong Kong. Results show that for France, prediction performance is similar, and for other countries, the regularized model outperforms the P-spline approach. Prediction errors are in the range of standard age-period-cohort models, except for the USA, where the P-spline model fails at predicting death rates. We investigate this point more in detail.

Table 4. Mean Absolute Error for USA, UK, France, Belgium (BEL), Japan (JAP) and Hong Kong (HKG) for the P-spline model in Currie et al. (2004) and our regularized model, based on historical data from 1986 to 2007. In bold: Lowest MAE per Country.

	Mean Absolute Error (10-Year Forecast)					
	USA	UK	France	BEL	JAP	HKG
P-splines	0.01405	0.00486	0.00232	0.00695	0.00384	0.00271
Regularized	0.00194	0.00212	0.00274	0.00244	0.00267	0.00214

5.3.1. The Case of the Usa

Figure 4 represents the logarithm of the crude death rates with the fitted and forecast death rates for USA both by the P-spline model and our regularized model from age 50 to 90. One striking observation is the erratic and unrealistic behaviour of the forecast death rates of the P-spline model: death rates around the age of 50 explode while death rates between 60 and 90 tend to converge to a single point. On the other hand, the elastic-net regularized model is able to capture properly the general trend of the mortality surface.

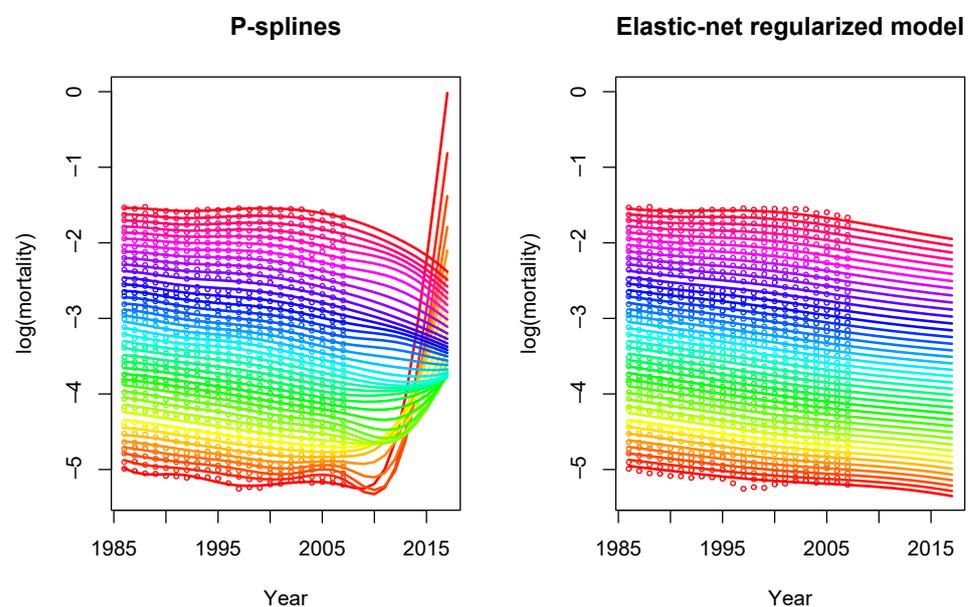


Figure 4. Logarithm of the fitted death rates and forecast death rates for USA by ages (50–90) over years (1986–2017). On the **left**: forecast by the P-spline model. On the **right**: forecast by our regularized elastic-net model.

The failure of the P-spline model to extrapolate death rates is due to the selection procedure of the smoothing parameters. We can observe on Figure 4 that the death rates at age 50 (bottom line on the figure) between 1995 and 2000 deviate from the general trend. Since smoothing parameters (λ_a, λ_y) aim at minimizing the BIC, and therefore the deviance, the P-spline model will select a low value of λ_y for the year dimension in order to

better fit these death rates. Indeed, the algorithm of the package **MortalitySmooth** chose $(\lambda_a, \lambda_y) = (0.01, 0.003)$ for a $BIC = 5270$. As a consequence, the regression coefficients at the age of 50 are not smoothed and present a lot of variability as shown on the right panel of Figure 5. The erratic behaviour of the coefficients at age 50 is then spread to adjacent coefficients and to the entire extrapolated mortality surface. One way to circumvent this issue is to increase the smoothing parameter on the year dimension λ_y , which will slightly decrease the goodness-of-fit but improve the extrapolation and the corresponding forecasts. The left panel of Figure 5 depicts the forecast of death rates in the USA with $\lambda_y = 20$ instead of $\lambda_y = 0.003$. For $\lambda_y = 20$, BIC slightly increases to 5335 but offers much more realistic forecasts.

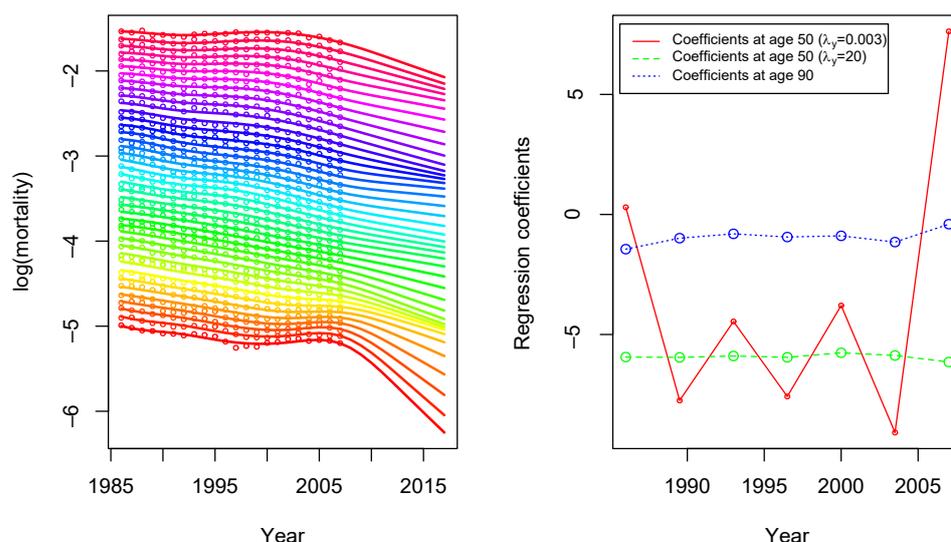


Figure 5. On the **left**: forecast by the P-spline model with $\lambda_y = 20$. On the **right**: regression coefficients in the P-spline model at age 50 with $\lambda_y = 0.003$ and $\lambda_y = 20$ and at age 90.

Our regularized model does not present this unrealistic behaviour because the set of parameters (λ, α) are selected by cross-validation with the presence of a lasso regularization term. Therefore, the fitted parameters will not be severely impacted if few points deviate from the general trend. The example of the USA highlights the importance of using out-of-sample techniques rather than an in-sample criterion such as the BIC.

5.3.2. The Case of COVID-19

In this section, we analyze how our regularized model and the P-spline model change their predictions by considering two scenarios: one optimistic and one pessimistic, as described in Section 4.4.

Since we have mortality data from HMD until 2017, we perturb the data as if there were a pandemic in 2013. As a first scenario, we assume that there is a death increase of 10% in 2013 for all ages, i.e. $\beta(x) = 0.1$, followed by a death decrease over the next two years (2014–2015), i.e., $K = 2$. The last two years (2016–2017) are not disturbed, assuming mortality reverts to normal. Figure 6 represents the logarithm of the crude death rates with the fitted and forecast death rates for France, both by the P-spline model and our regularized model after including the pandemic shock. As was the case for forecasting USA death rates, the P-spline provides unrealistic forecasts. Indeed, since death rates are perturbed for three years and then back to normal, we should expect a continuous decreasing trend. On the contrary, the P-spline predicts an increase of death rates at levels which are higher than the ones in 1986. On the other hand, the regularized model is robust towards the pandemic shock and predicts a decreasing trend as expected.

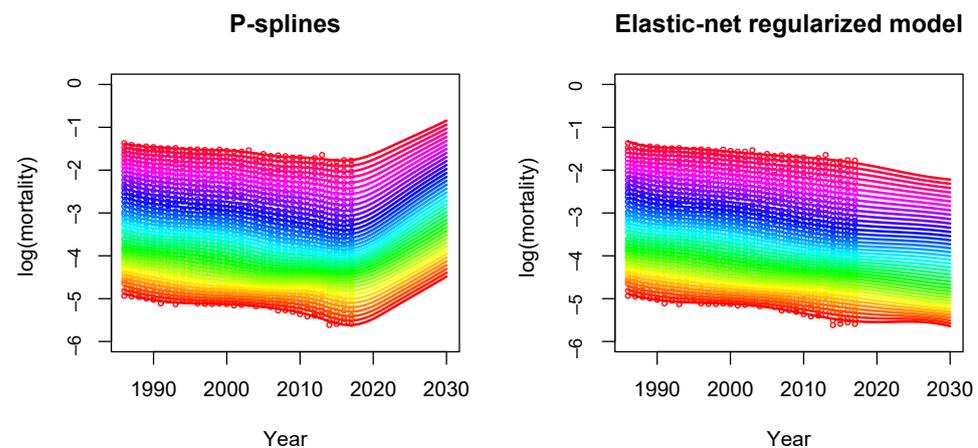


Figure 6. Logarithm of the fitted death rates and forecast death rates for France by ages (50–90) over years (1986–2030), after a pandemic shock in 2013 compensated by two years with lower mortality. On the **left**: forecast by the P-spline model. On the **right**: forecast by our regularized elastic-net model.

As a second scenario, we perturb mortality data for France as if there were a pandemic in 2013. In this case, we assume that there is a death increase of $\beta(x) = 10\%$ for all ages in 2013 and in the following years according to (9). Figure 7 represents the logarithm of the crude death rates with the fitted and forecast death rates, both by the P-spline model and our regularized model after including this second pandemic shock. Again, we observe that the P-spline model is unable to provide realistic forecasts. While we were expecting a mortality increase, the P-spline model predicts a mortality decrease and improbable mortality levels. By contrast, our regularized model is able to capture the pandemic shock and predicts a reduction in longevity improvement as expected. Overall, both pandemic scenarios considered here show that the regularized model provides adequate forecasts, which is not the case for the P-spline model.

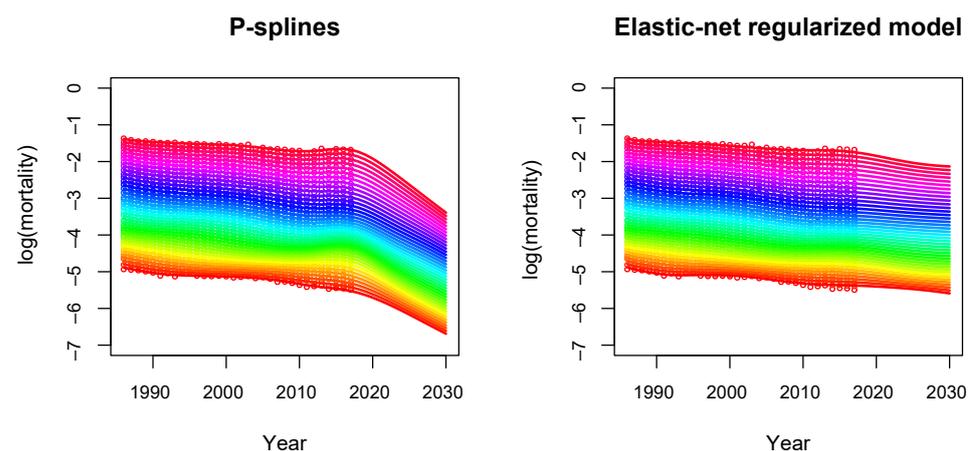


Figure 7. Logarithm of the fitted death rates and forecast death rates for France by ages (50–90) over years (1986–2030) after a pandemic shock in 2013, followed by a continuous mortality increase. On the **left**: forecast by the P-spline model. On the **right**: forecast by our regularized elastic-net model.

5.4. Model Uncertainty and Prediction Intervals

The extrapolation of the regularized mortality surface does not directly provide prediction intervals. In Section 4.5, we have introduced a simple procedure to obtain prediction intervals by the inclusion of a period index. These prediction intervals are centered around the central death rates estimated from the regularized elastic net model

and include the uncertainty from the period effect. To illustrate this procedure, Figure 8 depicts prediction intervals from the Lee–Carter model and from our regularized model for the French population. We observe that the general procedure produces central trend and uncertainty levels which are consistent with the ones produced by the Lee–Carter model.

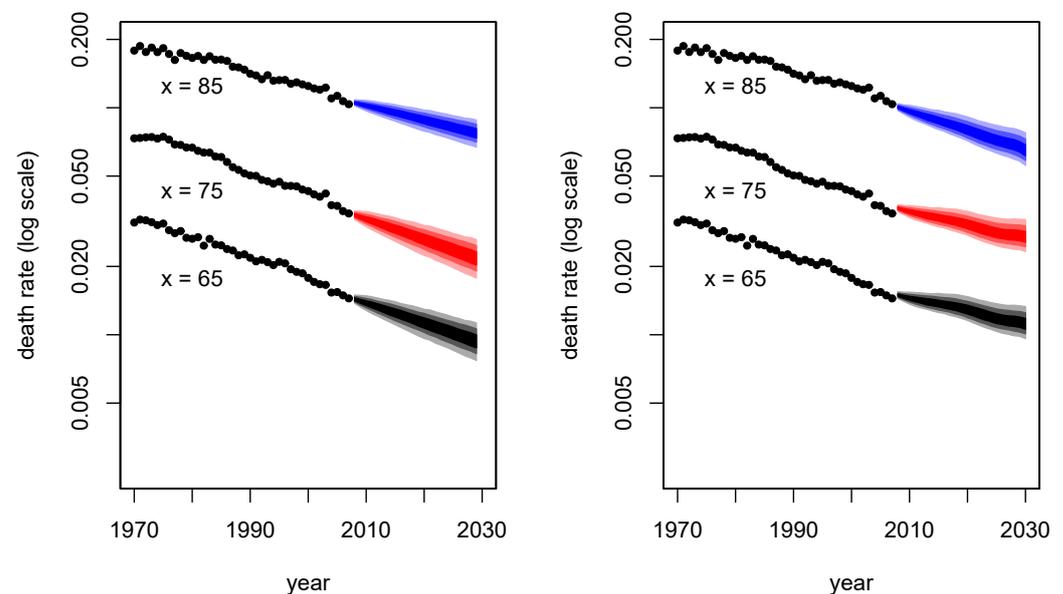


Figure 8. Fan charts for death rates $\mu_{x,t}$ at ages $x = 65, x = 75, x = 85$ from the Lee–Carter model on the left and from the regularized elastic net model on the right to the French population. The dots show historical death rates from 1970–2007 and shades in the fan represent prediction intervals at the 50%, 80% and 95% level.

6. Concluding Remarks

In this paper, we have proposed a new method to smooth mortality data in two dimensions, based on the elasticnet regularization and cross-validation. Comparisons with standard age-period-cohort models show that it provides robust estimates relative to the data used to fit the model and relative to the different countries considered and that the best model is far from being the same in different countries and zones (see, for instance, Tables 2 and 3). A detailed comparison with the P-spline model shows that the elasticnet regularized model is more robust to outliers, while the P-spline model can present erratic behaviour in presence of some data points that deviate from the general trend, like in the COVID-19 era. Indeed, in Section 5.3, we found that the P-spline approach leads to implausible forecasts for the American data and different data sets disturbed by pandemic shocks (see also Figures 6 and 7).

We believe that regularization techniques may help in the context of mortality modelling. Smoothness of the parameters in a stochastic mortality model is a key feature to obtain robust forecasts. As shown in this paper, parsimonious penalized models are less sensitive to the period of fit and their forecasts present less variability compared to models with cohort effects. Further research is needed concerning the application of regularization for the fitting and forecasting of mortality models. Moreover, recent work on regime-switching models (Gyls and Šiaulyš (2020)) shows promising results in modelling the effect of pandemic in the Swedish data. It would be interesting to consider a comparison between regime-switching models and smoothing models based on out-of-sample performance; this point is left for further research.

Author Contributions: Conceptualization, K.B., S.L. and Y.S.; methodology, K.B., S.L. and Y.S.; software, K.B., S.L. and Y.S.; validation, K.B., S.L. and Y.S.; formal analysis, K.B., S.L. and Y.S.; investigation, K.B., S.L. and Y.S.; resources, K.B., S.L. and Y.S.; data curation, K.B., S.L. and Y.S.; writing—original draft preparation, K.B., S.L. and Y.S.; writing—review and editing, K.B., S.L. and Y.S.; visualization, K.B., S.L. and Y.S.; supervision, K.B., S.L. and Y.S.; project administration, K.B., S.L. and Y.S.; funding acquisition, K.B., S.L. and Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work is mainly supported by the AXA research fund project on “Niveaux et tendances d’amélioration de la longévité: sélection de modèle pour les hypothèses “Best Estimate” et détection de rupture à l’aide de tests séquentiels, avec prise en compte du contexte covid-19”. S. Loisel and Y. Salhi also acknowledges support from the BNP Paribas Cardif Chair “Data Analytics & Models for Insurance” and the Milliman research initiative “Actuariat Durable”.

Acknowledgments: The authors thank the referees for their comments which helped improving the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Barrieu, Pauline, Harry Bensusan, Nicole El Karoui, Caroline Hillairet, Stéphane Loisel, Claudia Ravanelli, and Yahia Salhi. 2012. Understanding, modelling and managing longevity risk: Key issues and main challenges. *Scandinavian Actuarial Journal* 2012: 203–31. [CrossRef]
- Brouhns, Natacha, Michel Denuit, and Ingrid Van Keilegom. 2005. Bootstrapping the poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal* 2005: 212–24. [CrossRef]
- Cairns, Andrew J. G., David Blake, and Kevin Dowd. 2006. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance* 73: 687–718. [CrossRef]
- Cairns, Andrew J. G., David Blake, Kevin Dowd, Guy D. Coughlan, David Epstein, and Marwa Khalaf-Allah. 2011. Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics* 48: 355–67. [CrossRef]
- Cairns, Andrew J. G., David Blake, Kevin Dowd, Guy D. Coughlan, David Epstein, Alen Ong, and Igor Balevich. 2009. A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal* 13: 1–35. [CrossRef]
- Cairns, Andrew J. G., David P. Blake, Amy Kessler, and Marsha Kessler. 2020. The Impact of COVID-19 on Future Higher-Age Mortality. Available online: <https://ssrn.com/abstract=3606988> (accessed on 19 May 2020).
- Camarda, Carlo G. 2012. Mortalitysmooth: An r package for smoothing poisson counts with p-splines. *Journal of Statistical Software* 50: 1–24. [CrossRef]
- Currie, Iain D. 2016. On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal* 2016: 356–83. [CrossRef]
- Currie, Iain D, Maria Durban, and Paul HC Eilers. 2004. Smoothing and forecasting mortality rates. *Statistical Modelling* 4: 279–98. [CrossRef]
- Doukhan, Paul, Denys Pommeret, Joseph Rynkiewicz, and Yahia Salhi. 2017. A class of random field memory models for mortality forecasting. *Insurance: Mathematics and Economics* 77: 97–110. [CrossRef]
- Doukhan, Paul, Joseph Rynkiewicz, and Yahia Salhi. 2020. Optimal Neighborhoods Selection for AR-ARCH Random Fields with Application to Mortality. Available online: <https://hal.archives-ouvertes.fr/hal-02455803> (accessed on 1 November 2020).
- Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33: 1. [CrossRef]
- Guibert, Quentin, Olivier Lopez, and Pierrick Piette. 2019. Forecasting mortality rate improvements with a high-dimensional var. *Insurance: Mathematics and Economics* 88, 255–72. [CrossRef]
- Gylys, Rokas, and Jonas Šiaulyš. 2020. Estimation of uncertainty in mortality projections using state-space lee-carter model. *Mathematics* 8: 1053. [CrossRef]
- Haberman, Steven, and Arthur Renshaw. 2011. A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics* 48: 35–55. [CrossRef]
- Hainaut, Donatien. 2012. Multidimensional lee–Carter model with switching mortality processes. *Insurance: Mathematics and Economics* 50: 236–46. [CrossRef]
- Human Mortality Database. 2019. University of California, Berkeley (Usa) and Max Planck Institute for Demographic Research (Germany). Available online: www.mortality.org (accessed on 18 June 2020).
- Hunt, Andrew, and David Blake. 2014. A general procedure for constructing mortality models. *North American Actuarial Journal* 18: 116–38. [CrossRef]
- Hunt, Andrew, and Andrés M Villegas. 2015. Robustness and convergence in the lee–carter model with cohort effects. *Insurance: Mathematics and Economics* 64: 186–202. [CrossRef]

- Koissi, Marie-Claire, Arnold F Shapiro, and Göran Högnäs. 2006. Evaluating and extending the lee–Carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics* 38: 1–20. [[CrossRef](#)]
- Kyung, Minjung, Jeff Gill, Malay Ghosh, George Casella. 2010. Penalized regression, standard errors, and bayesian lassos. *Bayesian Analysis* 5: 369–411. [[CrossRef](#)]
- Lee, Ronald D., and Lawrence R. Carter. 1992. Modeling and forecasting us mortality. *Journal of the American Statistical Association* 87: 659–71.
- Lockhart, Richard, Jonathan Taylor, Ryan J. Tibshirani, and Robert Tibshirani. 2014. A significance test for the lasso. *Annals of Statistics* 42: 413. [[CrossRef](#)]
- Lovász, Enrico. 2011. Analysis of finnish and swedish mortality data with stochastic mortality models. *European Actuarial Journal* 1: 259–89. [[CrossRef](#)]
- Milidonis, Andreas, Yijia Lin, and Samuel H. Cox. 2011. Mortality regimes and pricing. *North American Actuarial Journal* 15: 266–89. [[CrossRef](#)]
- Millossovich, Pietro, Andrés M Villegas, and Vladimir K. Kaishev. 2018. Stmomo: An r package for stochastic mortality modelling. *Journal of Statistical Software* 84. [[CrossRef](#)]
- Pitacco, Ermanno, Michel Denuit, Steven Haberman, and Annamaria Olivieri. 2009. *Modelling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press: Oxford, UK.
- Renshaw, Arthur E., and Steven Haberman. 2006. A cohort-based extension to the lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* 38: 556–70. [[CrossRef](#)]
- Renshaw, Arthur E., and Steve Haberman. 2008. On simulation-based approaches to risk measurement in mortality with specific reference to poisson lee–Carter modelling. *Insurance: Mathematics and Economics* 42: 797–816. [[CrossRef](#)]
- Tibshirani, Ryan J., Alessandro Rinaldo, Rob Tibshirani, Larry Wasserman. 2018. Uniform asymptotic inference and the bootstrap after model selection. *The Annals of Statistics* 46: 1255–87. [[CrossRef](#)]
- Zou, Hui, and Trevor Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67: 301–20. [[CrossRef](#)]