Stockholms
universitet

Spline Models as Flexible Alternatives to the Lee-Carter Model

Filip Walldén

Masteruppsats i försäkringsmatematik
Master Thesis in Actuarial Mathematics

Masteruppsats 2026:15
Försäkringsmatematik
Juni 2026

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2026:15**
http://www.math.su.se

Spline Models as Flexible Alternatives to the Lee-Carter Model

Filip Wallldén*

June 2026

Abstract

One of the many challenges in insurance pricing is to accurately model and predict mortality, a task for which the Lee-Carter model has served as a benchmark in Sweden for several decades. In this study, the spline models will be examined as an alternative. Spline models are a set of flexible functions designed for interpolation and the creation of smooth surfaces. In total, two Lee-Carter models will be constructed: a standard model and a smoothed model, as commonly used in Swedish mortality models, as well as three spline models with different bases and penalties. Mortality rates will then be forecast, and the predicted mortality rates will be compared against the observed mortality rates for several data periods and metrics, including term life insurance and life annuity pricing errors. The spline models will prove to be less dependent on the training period than the Lee-Carter models, while the Lee-Carter models will be shown to estimate more accurate and stable pricing for modern data with aggregated population and death counts. For the gender-specific models, the spline models will provide marginally more accurate average pricing but will be more sensitive to the mortality shock observed during the COVID-19 period near the boundary.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: Filipwallden@hotmail.com. Supervisor: Filip Lindskog.

Acknowledgments

I would like to thank my supervisor, Filip Lindskog, for his guidance and patience, as well as my girlfriend, Jacqueline, for her feedback and assistance with language refinement. This thesis would not have been possible without their support.

AI Usage Statement

Artificial intelligence tools were used to assist with grammar and technical phrasing.

Contents

1 Introduction

An insurance company's primary purpose is to provide financial security to its customers, but the pricing of insurance products depends on future random events such as operational costs, claim severity, and inflation. In the case of life insurance, accurate estimation of mortality rates is of utmost importance. Underestimating mortality may lead to underpricing and, in extreme cases, leave the company unable to pay all claims. Overestimating mortality, on the other hand, may result in products becoming too expensive for customers. This makes mortality models with accurate forecasting abilities highly important, with the Lee-Carter model (Lee and Carter 1992) establishing itself as one of the dominant approaches over the past decades. The Lee-Carter model is stable, relatively simple, and has proven to be highly effective for mortality forecasting. However, the model is not perfect, as it can struggle with extreme ages and mortality shifts due to its inflexibility, although, for forecasting purposes this is not always impactful. In this study, we will examine an alternative to the Lee-Carter model: tensor product splines. The spline models are designed for interpolation and flexibility, and therefore do not struggle with mortality transitions or periods of shock in the same way as the Lee-Carter model; however, they have limited theoretical justification for extrapolation. Despite this, we find that the tensor product splines forecast mortality surprisingly well compared to the Lee-Carter models, while remaining less sensitive to the training period.

In Section 2, we will briefly present Swedish population and death count data from 1751 to 2022 and calculate necessary quantities for the models under consideration. In general, the dataset contains many periods of shock and mortality shifts, but largely stabilizes around the 1950s.

We will then present the necessary theory in Section 3, including the Lee-Carter model, splines, several computational optimizations, and the metrics used to compare the models. In this study, we will examine the P-spline and the thin plate spline. The P-spline is chosen because of its simple form and ease of use, while still enabling control over the flexibility of the smoothness, whereas the thin plate is a more ambitious spline basis designed to address a lot of the shortcomings of the common splines.

In Section 4, we will present, estimate, and compare the different models using out-of-time validation across several training lengths. This study will examine both aggregated models, fit using total population and death counts, and sex-specific models, in which males and females are modeled separately. We will find that the Lee-Carter model provides more stable pricing estimates than the spline models across both aggregated and sex-specific datasets, while generally producing more accurate pricing for the aggregated data. However, the differences are relatively small and largely depend on the age and gender of the insured individual. For the sex-specific models, the spline models often provide slightly more accurate pricing, although longer training periods are required for accurate estimation. The section concludes with a short study of mortality beyond 2022, including calculations of life expectancy for a 65-year-old individual and 50-year mortality forecasts for several age groups.

The results are summarized in Section 5 with a focus on the model differences in pricing and forecasting ability, emphasizing the stability of the Lee-Carter models and the potential of the spline-based alternatives.

We will close out the study with a brief discussion of the findings in Section 6 while highlighting a few areas where future development is possible.

2 Data

The data in this study were obtained from Human Mortality Database (2024) and consist of annual population-level mortality counts and corresponding exposures, aggregated by age and calendar year in Sweden for the period of 1751–2022. Each observation includes population and death counts reported both as totals and disaggregated by sex (female and male). The dataset is further grouped by ages, ranging from 0 to 110+. In total, the dataset consists of 20,175,458 deaths over 30,192 observations.

Assuming that deaths occur roughly uniform throughout the year, the central exposure to risk, $E^c_{x,t}$, can be approximated as

$$E^c_{x,t} = \frac{p_{x,t} + p_{x-1,t}}{2},$$

where $p_{x,t}$ is the population aged x at the beginning of year t. The special case of age 0 must be treated separately, with

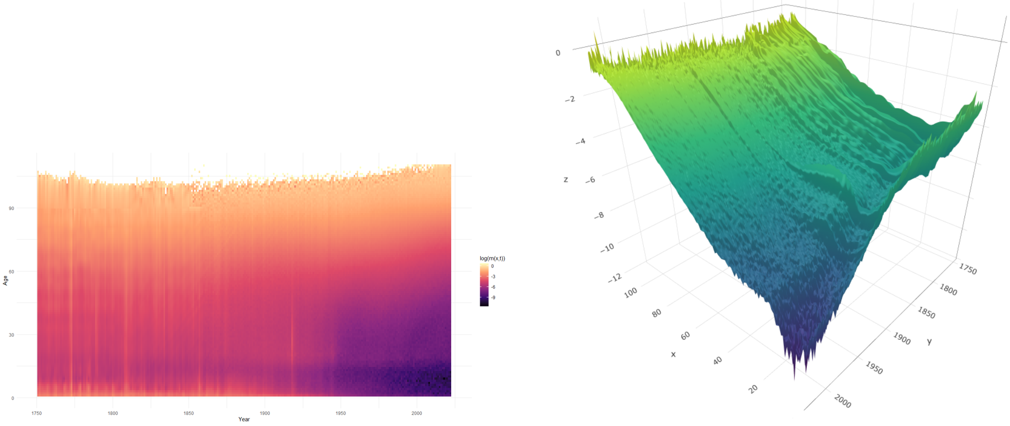
$$E^c_{0,t} = b_t,$$

where b_t is the number of births in year t. Due to the standard Lee-Carter model's (Lee and Carter 1992) reduced suitability at very young and very old ages (Booth et al. 2006), the data will be restricted to ages 10–90. Consequently, the special case above is only relevant to the data section.
The central death rate, $m_{x,t}$, can then be calculated as

$$m_{x,t} = \frac{d_{x,t}}{E^c_{x,t}},$$

where $d_{x,t}$ is the number of deaths between ages x and $x+1$ in year t. To prevent undefined central death rates, the data are further restricted by excluding observations with zero central exposure to risk.

Figures 1a and 1b plot the central death rates against year and age and show a sharp decline in infant mortality and a subtle positive trend to life expectancy in modern years. It is also possible to see patterns of instability with many years of mortality shifts, epidemics, and war, which the Lee-Carter model can struggle to accurately fit. Mortality later stabilizes in the 1950s with more recent years displaying a much smoother surface, which motivates the examination of smoothers.



(a) Heatmap of Log Mortality Rates (b) Mortality surface

Figure 1: Log mortality plotted against year and age

3 Theory

3.1 Lee-Carter

The Lee-Carter model (Lee and Carter 1992) is a simple but powerful statistical model for forecasting mortality rates and has become one of the standard models for mortality forecasting in Sweden (Svensk Försäkring 2021; Svensk Försäkring 2023). The model is defined as

$$\log(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t} \tag{1}$$

where $m_{x,t}$ is the mortality rate at age x and year t, a_x is the average mortality at age x across time, b_x describes how sensitive age x is to the time trend, k_t is a time series that captures the mortality trends, and $\epsilon_{x,t}$ is an error term. To ensure identifiability of the model parameters, the constraints

$$\sum_t k_t = 0 \qquad \text{and} \qquad \sum_x b_x = 1$$

are imposed. The simple form of the model makes extrapolating natural, as it only requires forecasting the time series k_t, which is often close to linear in time. However, the single parameterization of b_x can make it difficult for the model to capture mortality shifts or periods of shock, such as wars or pandemics. When considering more recent mortality data over shorter time spans, this tends not to be an issue, since mortality and standard of living have largely stabilized since the mid-1950s. Even so, the limitation motivates the search for more flexible mortality forecasting models.

3.2 Generalized Additive Models

The generalized linear models (GLMs) (Nelder and Wedderburn 1972) allow for non-normal response distributions and are written as

$$g(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$$

where $\mu = E(y)$, y is the response variable, $x_1, ..., x_n$ are a set of predictors, and g is what is referred to as a link function. The GLMs usually assume that y is from an exponential family distribution, which includes many useful distributions such as the Poisson distribution, which is often used to model the number of deaths over an interval. The generalized additive models (GAMs) (Hastie and Tibshirani 1986; Hastie and Tibshirani 1990) extend the GLMs to allow for smooth functions of the covariates and in their simplest form are written as

$$g(\mu) = \beta_0 + f_1(x_1) + f_2(x_2) + ... + f_n(x_n) \tag{2}$$

where f_i are smooth functions of the covariates x_i. Although not present in formula (2), the smooth functions can be extended to multiple covariates, allowing the construction of higher-dimensional smoothers, which are discussed in greater detail in Section 3.2.4.

3.2.1 Splines

The smooth functions, f_i, in formula (2) each consist of functions which, when summed, describe the full smooth function. That is

$$f_i(x) = \sum_{j=1}^{k} \beta_{i,j} b_{i,j}(x)$$

where $b_{i,j}$ is some known basis function, β_i is an unknown parameter, and k is the number of basis functions each smooth function is constructed with, also known as the number of knots. The choice of basis functions is of great importance, as they construct the smooth function and, to a large degree, determine its shape and properties. It is worth noting that the very common B-spline basis is only locally supported, with a large portion of each basis function's domain set to zero. In Figure 2, created by Currie, Durban, and Eilers (2004, p. 283), the construction of one such cubic B-spline with 8 knots is shown and illustrates how the individual basis functions build up the full smooth function. By following the dotted line at 1970, it is possible to see that the smooth function is

$$f(1970) = 0.0817\hat{\beta}_3 + 0.6267\hat{\beta}_4 + 0.2901\hat{\beta}_5 + 0.0016\hat{\beta}_6.$$

This enables the splines to become incredibly flexible, which leads to the question: how much flexibility is appropriate? If the spline is too smooth, the model will have difficulties fitting the data, but if it is too flexible, there is a risk of the function oscillating wildly and the model overfitting. To solve this, a 'wiggliness' penalty is

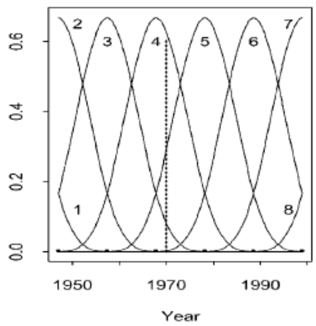


Figure 2: A basis with k = 8 highlighting the construction of splines, credit to Currie, Durban, and Eilers (2004, p. 283) for the plot

introduced to the fitting objective. Exactly what wiggliness means is determined by the user through the penalty function, which is a function that quantifies how wiggly the spline is, equating to 0 for perfectly smooth functions and to high values for wiggly functions. One common penalty function is

$$\int f''(x)^2 dx$$

which penalizes second-order derivatives. By defining a parameter λ that controls the tradeoff between smoothness and goodness of fit to the observed data, the model-fitting objective becomes a penalized regression problem in which

$$||\mathbf{y} - \mathbb{X}\boldsymbol{\beta}||^2 + \lambda \int f''(x)^2 dx \tag{3}$$

is minimized with respect to $\hat{\beta}$. This allows λ to control the degree to which the wiggliness should be penalized and is therefore called the smoothing parameter and treated as another parameter to be estimated. Methods for selecting λ, including restricted maximum likelihood (REML), are discussed in Section 3.3.

Splines are primarily used for interpolation and smoothing within data regions as the extrapolation behavior can be unpredictable and does not always reflect the full global trend in the data. This is because spline extrapolation proceeds in an unpenalized direction outside the observed data range, which under the common second-order derivative penalty results in linear continuation beyond the data boundary. As many spline models are constructed with local basis functions, extrapolation is primarily influenced by the basis functions near the boundary of the observed data, causing it to depend more on local boundary trends than on the global trend present in the data. This places great importance on the smoothing parameter in this study as the wiggliness directly influences the extrapolation behavior. This extrapolation problem, together with the work of Currie, Durban, and Eilers (2004) and Hall and Friel (2011), motivates the examination of two spline bases: the P-spline, which is simple while still allowing control over the penalty structure, and the more advanced thin plate spline. The former was applied to mortality modeling by Currie, Durban, and Eilers (2004), while the latter was considered by Hall and Friel (2011).

3.2.2 P-Spline

The B-splines were originally developed by Boor (1978), and have proven to be very useful as the basis functions are strictly local between the $m+3$ adjacent knots, where $m+1$ is the order of the basis. Eilers and Marx (1996) further developed what they called P-splines which are low-rank smoothers with B-spline basis defined on evenly spaced knots. In the case of splines, low rank refers to smoothers with a restricted number of basis functions, rather than one basis function per data point, thereby reducing the risk of overfitting and improving the computational efficiency in model estimation. Although the basis is able to handle uneven knot placement, one of the advantages of P-splines is their ease of use, which is unfortunately lost with unevenly placed knots. This is due to the fact that P-splines implicitly assume equal spacing between knots and apply difference penalties to the spline coefficients to control the wiggliness of the function. Consequently, uneven knot spacing requires additional adjustments to the penalty according to knot distances.
In the case of a second-order difference penalty, the penalty is

$$\mathcal{P} = \sum_{i=1}^{k-2} (\beta_{i+2} - 2\beta_{i+1} + \beta_i)^2.$$

The main appeal of the P-spline is its simplicity while still allowing for control over the level of smoothness, which, as mentioned earlier, is especially important in this setting because smoothness strongly affects the resulting forecasts.

3.2.3 Thin Plate Spline

A lot of the spline bases require knot selection, both in number and location, which introduces additional uncertainty for model selection, and most are only designed for smooth functions of one predictor variable. The thin plate developed by Duchon (1977) is an attempt at addressing those issues, creating knot-free bases with any number of predictor variables, which has been described as 'optimal' in some certain limited sense (Wood 2017, p. 215, sec 5.5.1). This is done by estimating a smooth function, $f(x)$, by finding $\hat{f}$ that minimizes

$$\sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda J_{m,d}(f) \tag{4}$$

where λ is a smoothing parameter, $J_{m,d}(f)$ is a penalty function for the wiggliness of f, and m and d are the order of the derivative penalty as well as the dimension of the spline.
The penalty function, $J_{m,d}(f)$, is defined as

$$J_{m,d}(f) = \int_{\mathbb{R}} \sum_{V_1+\ldots+V_d=m} \frac{m!}{V_1!\ldots V_d!} \left(\frac{\partial^m f}{\partial x_1^{V_1} \ldots \partial x_d^{V_d}} \right)^2 dx_1 \ldots dx_d.$$

As a concrete example, the penalty function for a two-dimensional thin plate spline with a second-order derivative penalty is

$$J_{2,2}(f) = \int\int \left(\frac{\partial^2 f}{\partial x_1^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x_1 x_2}\right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2}\right)^2 dx_1 dx_2.$$

Duchon (1977) further showed that the function minimizing equation (4) has the form

$$\hat{f}(\mathbf{x}) = \sum_i \delta_i \eta_{m,d}(\|\mathbf{x} - \mathbf{x}_i\|) + \sum_j \alpha_j \phi_j(\mathbf{x})$$

where δ and α are vectors of coefficients to be estimated, with δ having the additional constraint of $\sum_i \delta_i \phi_j(\mathbf{x}_i) = 0$, and where ϕ_i are the $\binom{m+d-1}{d}$ linearly independent polynomials in $\mathbb{R}^d$ with degree strictly less than m where $J_{m,d}$ is zero. In the case of a two-dimensional thin plate spline with a second-order derivative penalty, ϕ_i becomes $\phi_1(\mathbf{x}) = 1$, $\phi_2(\mathbf{x}) = x_1$, and $\phi_3(\mathbf{x}) = x_2$. Finally, the basis functions, η, are defined as

$$\eta_{m,d}(r) = \begin{cases} \dfrac{(-1)^{m+1+d/2}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} r^{2m-d}\log(r) & d \text{ even} \\ \dfrac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} r^{2m-d} & d \text{ odd} \end{cases}$$

which gets reduced to $\eta_{2,2}(r) = r^2 \log r$ for a two-dimensional thin plate spline with second-order derivative penalty.
Which consequently reduces the spline fitting problem to minimizing

$$\|\mathbf{y} - \mathbf{E}\boldsymbol{\delta} - \mathbf{T}\boldsymbol{\alpha}\|^2 + \lambda\boldsymbol{\delta}^\top\mathbf{E}\boldsymbol{\delta}$$

with respect to $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\delta}}$, where $\mathbf{E}_{i,j} \equiv \eta_{i,j}(\|\mathbf{x} - x_i\|)$ and $\mathbf{T}_{i,j} = \phi_j(x_i)$.

This construction of $\hat{f}$ requires no knot positions or basis functions to be specified, while still maintaining some control over the wiggliness by selection of the order of derivatives used in the penalty, and it is for this reason that the thin plate can be seen as the ideal smooth function. The thin plate spline is not without fault, as the estimation of a thin plate model is very computationally expensive, carrying as many unknown parameters as there are observations. Fortunately, Wood (2003) showed that low-rank approximations can be constructed through an eigen-decomposition of the spline basis matrix $\mathbf{E}$ by retaining only the eigenvectors associated with the largest eigenvalues, substantially reducing computational cost while retaining a close approximation to the thin plate spline.
Start by finding the eigen-decomposition $\mathbf{E} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$, where $\mathbf{D}$ is a diagonal matrix containing the eigenvalues of $\mathbf{E}$ ordered such that $|D_{i,i}| \geq |D_{i+1,i+1}|$, and the columns of $\mathbf{U}$ are the corresponding eigenvectors. Let $\mathbf{U}_k$ denote the matrix consisting of the first k columns of $\mathbf{U}$, and let $\mathbf{D}_k$ be the upper-left $k \times k$ submatrix of $\mathbf{D}$. Restricting the spline coefficients to the subspace spanned by these eigenvectors, let $\boldsymbol{\delta} = \mathbf{U}_k\boldsymbol{\delta}_k$. Next, find an orthogonal column basis $\mathbf{Z}_k$ such that $\mathbf{T}^\top\mathbf{U}_k\mathbf{Z}_k = 0$, ensuring that the identifiability constraints associated with the polynomial null space are satisfied. Finally, restrict $\boldsymbol{\delta}_k$ by setting $\boldsymbol{\delta}_k = \mathbf{Z}_k\tilde{\boldsymbol{\delta}}$. This reparameterization incorporates

the constraints directly into the model and allows the rank-k approximation to be obtained by finding $\tilde{\boldsymbol{\delta}}$ and $\boldsymbol{\alpha}$ that minimize

$$||\mathbf{y} - \mathbf{U}_k\mathbf{D}_k\mathbf{Z}_k\tilde{\boldsymbol{\delta}} - \mathbf{T}\boldsymbol{\alpha}||^2 + \lambda\tilde{\boldsymbol{\delta}}^\top\mathbf{Z}_k^\top\mathbf{D}_k\mathbf{Z}_k\tilde{\boldsymbol{\delta}}.$$

This approximation reduces the computational cost of the model estimation from $\mathcal{O}(n^3)$ to $\mathcal{O}(k^3)$. Although computing the eigen-decomposition of $\mathbf{E}$ is itself computationally expensive, the overall cost is substantially reduced when $n \gg k$. Consequently, the low-rank thin plate spline retains much of the flexibility and desired smoothing properties of the full thin plate spline while remaining computationally feasible for larger datasets.

3.2.4 Tensor-Product

Although not all splines were designed for smooth functions of higher dimensions, there are still methods of scaling them up. In particular, there are two main categories of higher-dimensional smoothers: the isotropic smoothers, which assume that the level of smoothness is the same in all directions, and the anisotropic tensor products, which allow for separate levels of smoothness in each direction. As it is not reasonable to assume that both age and period vary on the same scale, mortality cannot reasonably be modeled using isotropic smoothers, motivating the construction of tensor products by combining the separate basis functions for each covariate. In the two dimensional case, with smooth functions f_1 and f_2 with covariates x_1 and x_2, let

$$f_1(x_1) = \sum_{i=1}^{K_1} \alpha_i a_i(x_1)$$

and

$$f_2(x_2) = \sum_{j=1}^{K_2} \beta_j b_j(x_2).$$

The tensor product basis is then constructed as:

$$f(x_1, x_2) = \sum_{i=1}^{K_1}\sum_{j=1}^{K_2} \gamma_{i,j} a_i(x_1) b_j(x_2)$$

where $\gamma_{i,j} = \alpha_i\beta_j$. This results in a surface with different levels of smoothness and penalties in each direction but is unfortunately not scale-invariant, as different scales of the covariates, such as measurements in different units, result in different surfaces. The reason is that the penalty function in the fitting criteria depends on the scale of the covariates, which in turn changes how much wiggliness is penalized in each direction. Wood (2006) later proposed a solution to this problem. Let $f_{x_1|x_2}(x_1)$ denote $f_{x_1,x_2}(x_1, x_2)$ where x_2 is held constant, and define $f_{x_2|x_1}(x_2)$ similarly. By penalizing the marginal wiggliness of the smooth in each direction,

$$J(f_{x_1,x_2}) = \lambda_{x_1}\int_{x_2} J_{x_1}(f_{x_1|x_2})dx_2 + \lambda_{x_2}\int_{x_1} J_{x_2}(f_{x_2|x_1})dx_1,$$

where λ_{x_1} and λ_{x_2} are smoothing parameters controlling the degree of smoothness in each direction independently. This allows the marginal penalties to be weighted separately, making the resulting tensor-product smooth invariant to the relative scale of the covariates.

3.3 Restricted Maximum Likelihood

The estimation of the smoothing parameter λ can be approached in two ways: either by minimizing prediction error using criteria such as AIC or generalized cross-validation (GCV), or by treating the wiggliness as a random effect and estimating the smoothing parameter λ as a variance component. In this study, the latter was chosen as the prediction error alternative is prone to occasionally undersmooth and GCV is more likely to yield multiple minima with greater variance in λ estimates (Ruppert, Wand, and Carroll 2003; Reiss and Ogden 2009). As such, maximum likelihood (ML) or restricted maximum likelihood (REML) is preferable (Wood 2011). This choice is further motivated by realizing that spline extrapolation is entirely dependent on how the spline model reaches its boundary, as the model continues in the same unpenalised direction, which in this study results in linear extrapolation, as the models are penalized by second-order derivatives. This makes undersmoothing potentially undesirable as this can result in more variance in forecasting.

In the mixed-model representation of penalized splines (Ruppert, Wand, and Carroll 2003), the coefficients associated with the penalty are treated as random effects, allowing the smoothing parameter λ to be interpreted and estimated as a variance component using ML or REML methods. One of the problems with the ML estimate is that it tends to underestimate variance components and produce biased estimates, which motivated Patterson and Thompson (1971) to propose REML as a bias-reducing alternative. As underestimating the smoothing parameter results in undersmoothing and less stable extrapolation, REML is used for smoothing parameter estimation. The smoothing parameter is estimated by maximizing the restricted likelihood with respect to λ, where the REML criterion is given by

$$\begin{aligned}2l_r(\lambda) = &\log(\sum_j \lambda_j \mathbf{S}_j) - \log(\mathbf{X}^\top \mathbf{W}\mathbf{X} + \sum_j \lambda_j \mathbf{S}_j) \\ &- \frac{(y-\mathbf{X}\hat{\beta})^\top \mathbf{W}(y-\mathbf{X}\hat{\beta})}{\sigma^2} - (n-p)\log(\sigma^2)\end{aligned}$$

where y denotes the response variable, $\hat{\boldsymbol{\beta}}$ denotes the estimated coefficients, n is the number of observations, p is the number of unpenalized parameters, σ^2 is the residual variance, $\mathbf{X}$ is the design matrix, $\mathbf{W}$ is a diagonal weight matrix from the iteratively reweighted least squares (IRLS) algorithm, and $\mathbf{S}_j$ is the penalty matrix for the j-th smooth term.

3.4 GAMs for Large Datasets

Recall the penalized regression problem (3) in Section 3.2.1:

$$||\mathbf{y} - \mathbb{X}\boldsymbol{\beta}||^2 + \lambda \int f''(x)^2 dx.$$

To estimate $\hat{\beta}$, the penalized iteratively reweighted least squares (PIRLS) algorithm is used to maximize the penalized likelihood, which requires calculation of the product $\mathbf{X}^\top \mathbf{W} \mathbf{X}$ where $\mathbf{X}$ is the design matrix and $\mathbf{W}$ is a diagonal matrix. If n is the number of observations and k is the number of basis functions, then $\mathbf{X}$ is an $n \times k$ matrix, which makes the memory and computational costs of calculating the matrix product very expensive and makes the fitting algorithm very time-consuming. Consequently, while the generalized additive models are very powerful for modeling data, they tend to struggle with large datasets and model fitting can quickly become unfeasible. This is also true for the dataset used in this study, as 31,912 observations over 272 years crashed `R` due to insufficient memory after 12 hours of computational time, and even a smaller model with a minimal number of knots fitted to data from 2000–2022 requires several minutes to converge. It is evident that another approach is necessary to model the full dataset, and Wood, Goude, and Shaw (2015) proposes a solution to this problem, which is implemented in the `mgcv` (Wood 2023) package in `R` through the `bam` function. The idea proposed in the paper is to calculate the chunks of $\mathbf{X}^\top \mathbf{W} \mathbf{X}$ incrementally rather than calculating the full product all at once, which prevents the storage of the full design matrix $\mathbf{X}$. Once a chunk has been calculated and used, it can be safely discarded to reduce memory usage. As the chunks are calculated separately, `bam` introduces additional optimizations by allowing multiple cores of the computer to calculate the chunks simultaneously via the `nthreads` argument. One further optimization is the implementation of the fast restricted maximum likelihood algorithm (fREML) by allowing REML to reuse the chunks, as the product is also necessary for estimation of λ. In full, the aforementioned optimizations significantly reduce computation time and storage requirements for model fitting, resulting in the previously unfeasible model being able to be fitted in just a few seconds.

3.5 Overdispersed Poisson Distribution

The Poisson distribution is used for modeling the number of times a random event occurs within a fixed time interval, and consequently, it is the natural distribution candidate for the mortality models considered in this study. However, the Poisson distribution assumes that the variance and mean are equal,

$$\mathrm{Var}(D_{x,t}) = \mathbb{E}(D_{x,t}),$$

which does not always hold for real population data, as there are often unexplained effects such as heterogeneity or temporal shock. These unexplained effects often introduce additional variance and cause the variance to exceed the mean. This is called overdispersion and can be modeled using the overdispersed Poisson distribution, which introduces a dispersion parameter $\phi > 1$ such that

$$\mathrm{Var}(D_{x,t}) = \phi \mathbb{E}(D_{x,t}).$$

3.6 Metrics

Since this study aims to compare the considered mortality models in an actuarial setting, the models are evaluated using out-of-time validation, where fitting is based

on observations up to year T and forecasts are compared against the empirical observations for $t > T$. This comparison consists of Poisson deviance, root mean squared error (RMSE), mean absolute error (MAE), and pricing errors. For the Poisson deviance, RMSE, and MAE, the comparisons will be calculated using the observed and estimated death counts.
While all metrics are included in the analysis, particular emphasis is placed on pricing error due to the insurance-focused nature of the study.

The Poisson deviance is a goodness-of-fit measure for models estimated under a Poisson distribution. For observed death counts y_i and model-predicted death counts $\hat{y}_i$, where $\hat{y}_i$ denotes the estimated Poisson mean $\mathbb{E}[Y_i]$, the Poisson deviance is defined as

$$D = 2\sum_{i=1}^{n}\left(y_i \log\left(\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i)\right).$$

To improve readability, this metric is scaled down by the number of observations throughout this study.
For observed death counts, y_i, and model-predicted death counts, $\hat{y}_i$, the RMSE and MAE are

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|.$$

Two common life insurance products are term life insurance policies and life annuities. Term life insurance is a product in which a benefit is paid upon death, whereas a life annuity provides payment at the end of each year as long as the insured remains alive. This means that term life insurance pays out in the event of death, and life annuities in the event of survival. Both products usually have a term length for which they provide security, with 20 years being very common and used throughout the analysis.

The expected present value of an n-year term life insurance contract is calculated with the probabilities of surviving k years followed by death in year $k+1$, while discounting each year by a fixed interest rate. Let p_x be the probability that a person aged x survives one year and q_x the probability that a person aged x dies within one year. Assuming a uniform distribution of deaths, these probabilities can be obtained from the central death rate m_x as

$$q_x = \frac{m_x}{1 + \frac{m_x}{2}}, \qquad p_x = 1 - q_x.$$

Further, let ${}_kp_x = p_x p_{x+1} \cdots p_{x+k-1}$ be the probability that a person aged x survives k years. Assuming payment of 1 krona in the event of death, it follows that the expected value, measured in kronor, of an n-year term life insurance policy, $A_{x:\overline{n}|}$, can be written as

$$A_{x:\overline{n}|} = vq_x + v^2 p_x q_{x+1} + v^3 {}_2p_x q_{x+2} ... + v^n {}_{n-1}p_x q_{x+n-1}$$
$$= \sum_{k=0}^{n-1} v^{k+1} {}_kp_x \, q_{x+k}$$

where v is the discount factor $\frac{1}{1+i}$, with the interest rate, i, set to 2% for this study.

To calculate the expected present value of an n-year life annuity, only the probabilities of surviving k years are necessary, while discounting each year by a fixed interest rate. Assuming payment of 1 krona at the end of each survival year, the expected present value, measured in kronor, of an n-year life annuity, $a_{x:\overline{n}|}$, can be written as

$$a_{x:\overline{n}|} = vp_x + v^2 {}_2p_x + ... + v^n {}_np_x$$
$$= \sum_{k=1}^{n} v^k {}_kp_x.$$

The predicted present values can then be compared against the corresponding present values calculated using the mortality rates observed in the validation period, denoted by $A^{\text{obs}}_{x:\overline{n}|}$ and $a^{\text{obs}}_{x:\overline{n}|}$. To assess forecasting accuracy, relative errors are calculated as

$$\text{RelErr}^A_{x,n} = \frac{\hat{A}_{x:\overline{n}|} - A^{\text{obs}}_{x:\overline{n}|}}{A^{\text{obs}}_{x:\overline{n}|}}$$

and

$$\text{RelErr}^a_{x,n} = \frac{\hat{a}_{x:\overline{n}|} - a^{\text{obs}}_{x:\overline{n}|}}{a^{\text{obs}}_{x:\overline{n}|}}.$$

4 Analysis

The purpose of this study is to investigate the accuracy of the Lee-Carter and spline models' mortality forecasts. Consequently, the theory for Lee-Carter and spline models was established in Sections 3.1 and 3.2.1, respectively. This section applies the established theory and compares the two approaches with the metrics described in Section 3.6 using out-of-time validation. Throughout this section, the Poisson deviance, root mean squared error (RMSE), and mean absolute error (MAE) metrics will be used to compare the observed and predicted death counts.

4.1 Model Specifications

This study will examine several Lee-Carter and spline models for a deeper understanding of the differences between the two approaches.
One Lee-Carter model will be fitted as described in Section 3.1, with another Lee-Carter model fitted with smoothed parameters as used by Svensk Försäkring (2021).
In this study, two different splines are considered: the P-spline and the thin plate

spline, as described in Sections 3.2.2 and 3.2.3, with an additional P-spline with penalty and knots in accordance with Currie, Durban, and Eilers (2004) and Continuous Mortality Investigation (2011).
In total, five mortality models will be considered, two Lee-Carter models and three spline models, with the code related to model fitting attached in Appendix A.

Lee-Carter

The standard Lee-Carter model, LC, models the log mortality rates and is fitted with the `StMoMo` (Villegas, Kaishev, and Millossovich 2023) package in `R` with no additional modifications.

The smoothed Lee-Carter model, LC_s, is estimated by extracting the parameters b_x and k_t, as seen in equation (1) in Section 3.1, from the already fitted standard Lee-Carter model. The age-specific sensitivity parameter b_x is smoothed with a 5-year moving average, while the time index k_t is approximated with a linear equation by least squares, after which the model is reconstructed using the smoothed parameters to obtain the smoothed Lee-Carter model.

Spline

The spline models are fitted to death counts using the `bam` function from the `mgcv` (Wood 2023) package in `R`, which includes all the aforementioned optimizations, including the scale-invariant tensor product modifications, and the large dataset approximations discussed in Sections 3.2.4 and 3.4. The thin plate spline, S_{tp}, is fitted as a one-dimensional spline of `Age` and a two-dimensional tensor product of `Age` and `Year` with an offset of $\texttt{log(}E^c_{x,t}\texttt{)}$. The one-dimensional and two-dimensional splines were previously discussed in Sections 3.2.1 and 3.2.4 and are specified using `s()` and `te()`, respectively, in `mgcv`. The full model formula is thus written as

$$\texttt{deaths} \sim \texttt{s(Age) + te(Age,Year) + log(}E^c_{x,t}\texttt{)}.$$

The P-spline, S_{ps}, is specified with the same formula, while the Currie P-spline, S_c, drops the one-dimensional spline and is fit with only the tensor product as

$$\texttt{deaths} \sim \texttt{te(Age,Year) + log(}E^c_{x,t}\texttt{)}.$$

Both P-splines are fitted with cubic bases and second-degree penalties. For the Currie P-spline, a minimum penalty of 10^{-4} is applied to both `Age` and `Year`, as recommended by Continuous Mortality Investigation (2011), although no starting penalty is implemented as `bam` does not support it. This should only impact computational time and memory storage as REML tends to not produce multiple local maxima.
As insignificant knots and effective degrees of freedom are reduced through the smoothing parameter, the choice of the number of knots is largely arbitrary. It is therefore only necessary to specify a sufficiently large number of knots to ensure adequate model flexibility, while still maintaining computational feasibility. For this purpose, the number of knots is chosen as one knot every 4 years, with a maximum

of 20 knots in both directions, `Age` and `Year`, for all splines. The exception is the one-dimensional spline, `s(Age)`, in the thin plate spline model, where the number of knots is set to 80. All three spline models are fitted assuming an overdispersed Poisson distribution, as described in Section 3.5, as models fitted assuming a Poisson distribution result in overdispersion of 1.3–8 depending on the model and training period.

To calculate future predictions, the Lee-Carter model uses `fitted(lc_fit, type = "rates")` to return mortality rates, and the spline models use `predict(spline_fit, type = "response")` to return death counts. Conversion between mortality rates and death counts is done by either dividing or multiplying by corresponding exposure, and which one is preferred depends on the metric under consideration.

4.2 Training Period

As the dataset contains 272 years of death counts and exposures, and it is not immediately obvious how each model reacts to training periods of different lengths, it is necessary to examine models with different training lengths to establish a baseline training length with minimal model bias for future analysis. This is performed by training the models with datasets of varying lengths up to the year 2000 and comparing the forecasted mortalities for the next 20 years against the observed mortalities using the previously established metrics in Section 3.6. In this study, five different training lengths will be considered: 10, 20, 30, 50 and 250 years.

For this section, pricing errors are evaluated for a 20-year term life insurance policy issued to a 40-year-old and a 20-year life annuity issued to a 60-year-old, as these are among the most common ages at which such products are purchased. This analysis is repeated for both aggregated mortality models and models separated by sex.

4.2.1 Aggregated Mortality

1751–2000

Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	35.61	315.75	168.26	**-1.62**	-14.7
LC_s	257.13	835.87	548.81	78.98	-64.44
S_{tp}	**3.46**	**58.95**	**36.42**	3.07	-4.19
S_{ps}	5.56	83.54	43.03	4.21	-2.14
S_c	5.94	89.3	45.22	2.95	**-2.12**

Table 1: Forecasting performance for aggregated mortality during 2001–2020 with 250 years of training data

1991–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	6.64	80.02	45.03	**-1.05**	-2.37
LC_s	**4.77**	**61.24**	**36.28**	-1.15	-1.92
S_{tp}	5.54	102.24	50.06	2.72	**-1.41**
S_{ps}	5.77	109.83	54.78	-1.96	-4.27
S_c	6.89	93.71	53.25	-3.09	-3.81

Table 2: Forecasting performance for aggregated mortality during 2001–2020 with 10 years of training data

Tables 1, 2, and the tables in Appendix B contain the metrics for forecasting performance for aggregated mortality models. In these tables, it is possible to see that the spline models are able to accurately predict mortality for 2001–2020 across all training periods, both in terms of statistical fit and insurance pricing, but that the Lee-Carter models struggle with longer training periods. The smoothed Lee-Carter model is especially unsuited for forecasting mortality, as the linear time trend is not appropriate for the full dataset, as can be seen in Figure 3. In terms of statistical fit,

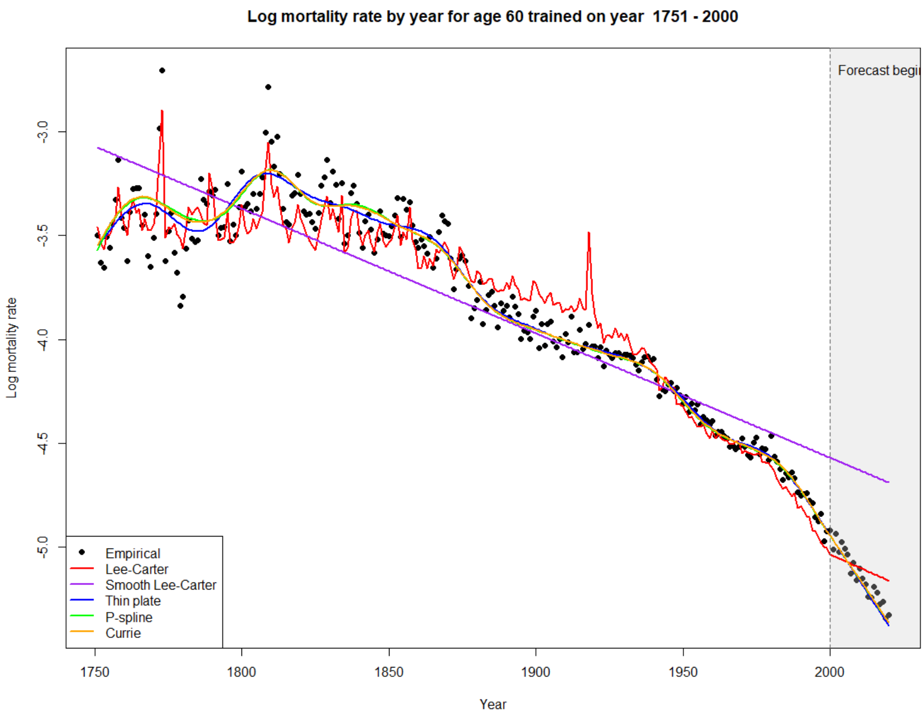


Figure 3: Comparison of the different models trained on 1751–2000 data and the empirical log mortality rate at age 60, with forecasts up to 2020

the Lee-Carter models perform best with 20–30 years of training data. However, as the primary focus of this study is for insurance purposes, the aggregated mortality models will subsequently be fitted using 10 years of training data, as this produces more accurate pricing forecasts across models without any obvious model bias.

4.2.2 Male Mortality

1981–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	**4.57**	**60.66**	**34.37**	-4.78	-9.49
LC_s	4.94	67.87	38.61	**-1.49**	-12.33
S_{tp}	6.15	71.51	39.48	-4.87	-7.77
S_{ps}	10.29	83.72	47.42	-3.61	**-5.03**
S_c	6.85	70.63	39.71	-6.87	-6.1

Table 3: Forecasting performance for male mortality during 2001–2020 with 20 years of training data

1991–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	6.92	55.15	34.25	-10.1	-5.24
LC_s	4.8	**48.82**	**30.08**	-8.15	-6.4
S_{tp}	**4.39**	53.48	33.45	-10.92	-5.5
S_{ps}	9.18	76.56	45.67	-13.56	7.02
S_c	5.7	49.78	32.24	**-6.83**	**-4.43**

Table 4: Forecasting performance for male mortality during 2001–2020 with 10 years of training data

As can be seen in Tables 3, 4, and the tables in Appendix B, the male mortality models generally favored training lengths of 10–20 years. Ten years of training data minimizes annuity pricing error, but leads to comparatively large errors for term life insurance products, whereas 20 years minimizes term life pricing error while increasing annuity pricing error. This suggests that the models trained on shorter training data more accurately predict old-age mortality, but less accurately predict middle-age mortality.

4.2.3 Female Mortality

1971–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	**2.38**	**34.31**	**20.43**	14.69	**-0.39**
LC_s	2.54	40.6	22.53	**12.39**	1.07
S_{tp}	3.7	55.63	29.4	14.4	0.42
S_{ps}	7.55	101.9	46.48	13.34	-0.7
S_c	3.23	50.33	26.07	14.31	0.76

Table 5: Forecasting performance for female mortality during 2001–2020 with 30 years of training data

1981–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	2.83	37	21.87	14.27	-0.28
LC_s	2.46	35.9	20.78	**14.14**	**-0.06**
S_{tp}	4.33	69.96	33.16	15.42	-1.44
S_{ps}	**2.2**	**32.42**	**19.03**	14.34	0.11
S_c	10.71	126.86	55.67	17.2	-1.07

Table 6: Forecasting performance for female mortality during 2001–2020 with 20 years of training data

Tables 5, 6, and the tables in Appendix B show that the female mortality models favor longer training periods than the male mortality models, as the 10-year training period produces exceptionally large annuity pricing errors. The optimal training length is 20–30 years, with a slight preference for 30 years due to slightly lower pricing errors compared to the 20-year models.

Both the male and female models largely agree with the aggregated results that the forecasts of the spline models are less training-period dependent and provide a closer fit to the full dataset. It is also clear that the sex-specific models require longer training periods than the aggregated models, most likely due to the additional variance introduced by the reduced population sizes.
As 20 years of training data was the only training length that provided accurate fits for both male and female models simultaneously without any obvious model bias, future sex-specific models will be fitted with 20 years of training data.

4.3 Pricing Accuracy

To compare the models' pricing accuracy, the models are fitted up to the year 2000 using the preferred training lengths, and the present values of 20-year term

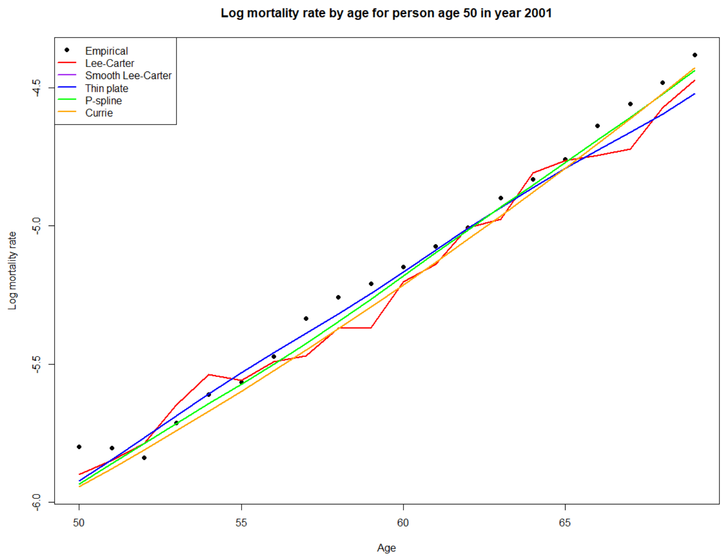


Figure 4: Log mortality rate for an individual aged 50 in 2001

life insurance contracts and 20-year life annuities issued in 2001 are calculated for different ages and compared against the empirical present values, as discussed in Section 3.6.

4.3.1 Aggregated Pricing

	Term Error (%)			
Model	40	50	60	70
LC	**-1.05**	-5.36	1.04	1.07
LC_s	-1.15	-5.72	0.96	**0.66**
S_{tp}	2.72	-4.42	**0.93**	3.22
S_{ps}	-1.96	**-3.62**	5.25	4.08
S_c	-3.09	-5.28	5.57	3.33

Table 7: Pricing errors for aggregated 20-year term life insurance policies by issue age for policies issued in 2001 using 10 years of training data

Table 7 shows the term life insurance pricing errors of the aggregated mortality models and contains several noteworthy patterns. Most importantly, the Lee-Carter models exhibit considerable stability, with pricing errors close to 1% across all ages except at age 50. All models produce inaccurate pricing at age 50, and all predictions are negative, suggesting that the models underprice the product and that the observed mortality is higher than predicted. This is confirmed by Figure 4. The plot shows the predicted and observed log mortality rates of a person aged 50 in 2001, and it can be seen that the models underestimate mortality at ages 55–69 in the years 2006–2020.

Table 7 also shows that the underestimation of mortality is present to a lesser extent at age 40 for all models, with the exception of the thin plate spline, and that there is a slight overestimation of mortality for all models at older ages. Overall, the Lee-Carter models provide the most accurate pricing of 20-year term life insurance products for aggregated mortality, with the standard Lee-Carter model fitting younger ages more accurately and the smoothed Lee-Carter model fitting older ages more accurately. The P-spline produces the lowest pricing errors among the spline models for younger and middle ages, but performs poorly at older ages, whereas the thin plate spline yields slightly less accurate predictions at younger and middle ages and substantially more accurate predictions at older ages. The Currie P-spline struggles compared to the other models, producing inaccurate pricing across all ages.

Annuity Error (‰)				
Model	40	50	60	70
LC	-0.48	1.52	-2.37	-2.68
LC_s	-0.4	1.66	-1.92	**-1.92**
S_{tp}	-0.63	**0.89**	**-1.41**	-5.16
S_{ps}	**-0.13**	1.48	-4.27	-10.91
S_c	-0.23	2.38	-3.81	-11.58

Table 8: Pricing errors for aggregated 20-year life annuities across issue ages for policies issued in 2001 using 10 years of training data

The life annuity price estimates for the aggregated models in Table 8 are very similar across most models and ages, producing less than 1% pricing errors. This is to be expected as life annuities compound survival rates and produce natural averages in pricing. In contrast, the term life insurance prices depend on local mortality probabilities, making them sensitive to even smaller inaccuracies in mortality predictions. Life annuity prices are generally higher than those of term life insurance products, making pricing inaccuracies smaller relative to the total price.

As shown in the table for term life pricing errors, mortality at older ages is overestimated, which is seen through underpriced life annuities. Both products further agree that the mortality at age 50 is underestimated. It is only at age 40 that the two insurance products suggest different mortality biases, with the life annuity pricing estimates indicating an overestimation of mortality and the term life insurance pricing estimates indicating an underestimation. It might not be immediately clear how a term life insurance contract, which pays upon death, and a life annuity, which pays upon survival, can both be underpriced. Figure 5 helps illustrate how this is possible. The plot shows an overestimation of mortality in the early years of the insurance, which results in longer life annuities and more payouts than expected. Simultaneously, there is a slight underestimation of mortality in later years, which results in more payouts for term life insurance contracts. This overestimation of mortality in the early years and underestimation in later years results in more payouts than initially predicted for both products.

Table 8 further shows the stability of the Lee-Carter models' forecasts, which price older ages with less than 0.3% pricing error, with the smoothed Lee-Carter model

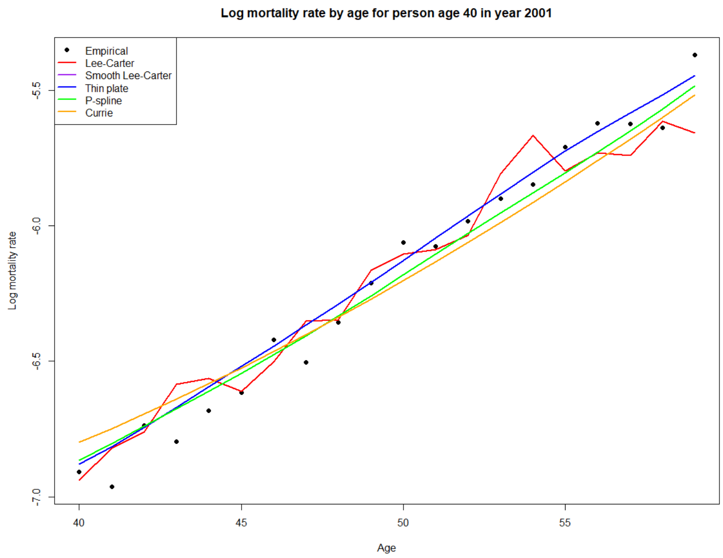


Figure 5: Log mortality rate for a 40-year-old person in 2001

providing the most accurate forecasts. The spline models are also very accurate: the P-splines are exceptional at age 40, while the thin plate spline produces the most accurate pricing at ages 50 and 60. All three spline models are unable to accurately forecast life annuities at age 70, and as annuities are common at older ages, the overall best spline model would be the thin plate, as it produces the most accurate predictions for ages 50 and over.

It is clear that whilst the Lee-Carter models, particularly the smoothed variant, provide the most accurate overall pricing estimates, the spline models also perform well for younger and middle ages.

4.3.2 Male Pricing

Term Error (%)				
Model	40	50	60	70
LC	-4.78	-3.95	8.22	**3.55**
LC_s	**-1.49**	**-1.09**	10.56	4.28
S_{tp}	-4.87	-7.91	4.93	4.22
S_{ps}	-3.61	-9.18	6.39	5.54
S_c	-6.87	-10.11	**3.6**	4.29

Table 9: Pricing errors for male 20-year term life insurance policies by issue age for policies issued in 2001 using 20 years of training data

The term life insurance pricing errors for the male mortality models in Table 9 show an underestimation of mortality at younger ages and an overestimation at older ages.

The Lee-Carter models are unable to accurately predict mortality at age 60, and the spline models produce poor predictions at age 50. Overall, the smoothed Lee-Carter model forecasts the most accurate term life pricing for younger ages and is therefore the most suitable model for pricing male term life insurance policies. The remaining models produce similar pricing errors, with the preferred model depending entirely on age.

Annuity Error (‰)				
Model	40	50	60	70
LC	**0.01**	2.44	-9.49	-16.97
LC_s	-0.49	**1.13**	-12.33	-22.28
S_{tp}	0.07	3.56	-7.77	-14.21
S_{ps}	0.65	2.88	**-5.03**	-18.15
S_c	0.09	4.66	-6.1	**-11.85**

Table 10: Pricing errors for male 20-year life annuities across issue ages for policies issued in 2001 using 20 years of training data

As with the term life pricing, the pricing errors for life annuities in Table 10 suggest that male mortality is underestimated at younger ages and overestimated at older ages. There is no clear best model for male annuities, as all models perform poorly at older ages where life annuities are most common, particularly the Lee-Carter models. However, the two strongest candidates are the P-spline and the Currie P-spline. Both models produce the lowest pricing errors at ages 60 and 70, with a slight edge to the Currie P-spline because of the standard P-spline's poor pricing forecast at age 70. The thin plate spline also estimates the prices well, with more accurate pricing for younger ages than the Currie P-spline, but less accurate pricing for older ages.

4.3.3 Female Pricing

Term Error (%)				
Model	40	50	60	70
LC	14.27	2.63	-1.39	-1.14
LC_s	**14.14**	1.95	-1.18	-1.28
S_{tp}	15.42	4.14	2.9	5.24
S_{ps}	14.34	**1.53**	**-1.16**	**-0.91**
S_c	17.2	4.65	4.56	10.37

Table 11: Pricing errors for female 20-year term life insurance policies by issue age for policies issued in 2001 using 20 years of training data

The female term life pricing errors in Table 11 show that younger ages are overpriced across all models, suggesting that female mortality at younger ages is overestimated

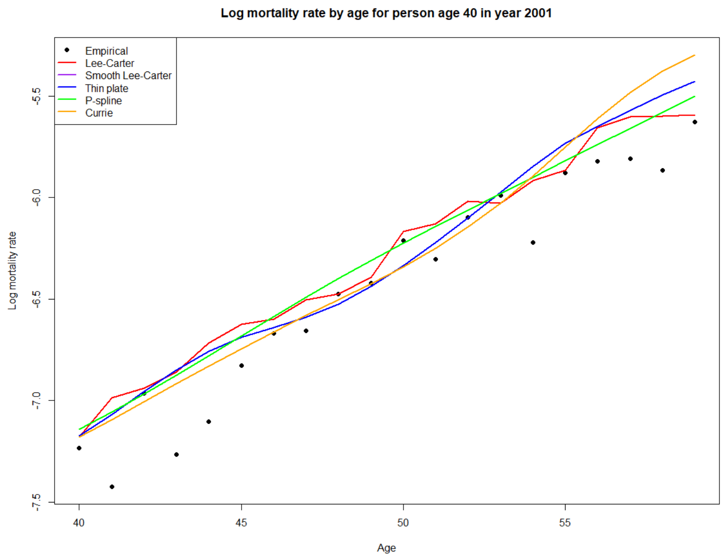


Figure 6: Log mortality rate for a 40-year-old woman in 2001

by the models. However, there is no such consensus for older ages, with some models suggesting overestimation of mortality and some suggesting underestimation. The predictions at ages 50, 60, and 70 are all very accurate, but it is clear that the forecasts at age 40 are very poor. This is illustrated by Figure 6, which shows a substantial overestimation of mortality for a 40-year-old woman purchasing a 20-year term life insurance policy in 2001.
Out of all the models, the P-spline provides the most accurate overall pricing, providing the most accurate pricing predictions at ages 50, 60, and 70, and is only 1.4% less accurate at age 40 than the smoothed Lee-Carter model, which produces the most accurate estimate for that age. Both Lee-Carter models produce accurate pricing forecasts as well, with pricing errors only slightly worse than the P-spline. Neither the thin plate nor the Currie P-spline is advised, as these produce significantly less accurate pricing across all ages, particularly at ages 60 and 70.

Annuity Error (‰)

Model	40	50	60	70
LC	-1.74	-1.5	-0.28	**4.06**
LC_s	-1.7	-1.26	**-0.06**	4.33
S_{tp}	-1.33	-1.73	-1.44	-7.27
S_{ps}	-1.66	**-1.11**	0.11	4.12
S_c	**-1.08**	-2.29	-1.07	-16.29

Table 12: Pricing errors for female 20-year life annuities across issue ages for policies issued in 2001 using 20 years of training data

The annuity pricing errors in Table 12 are similar across the two Lee-Carter models and the P-spline, with the preferred model depending on the age of the insured. The

P-spline provides the most accurate pricing predictions out of the three for younger ages and is only slightly worse for older ages, while the smoothed Lee-Carter is favored at age 60 and the standard Lee-Carter at age 70. Although the thin plate and Currie P-spline produce accurate pricing for younger ages, these models estimate poorly for older ages compared to the other three models.

For the sex-specific models, the male mortality models forecast younger ages more accurately than the female models, while the female models provided more accurate forecasts for older ages. The Lee-Carter model estimates are much less accurate for the sex-specific models, and the spline models are preferred for pricing. The two best spline models were the thin plate and P-spline, both of which estimated prices well, with the P-spline providing marginally more accurate pricing estimates.

4.3.4 Life Expectancy Estimates

It is also possible to use the models to calculate the life expectancy at age 65, although it is necessary to temporarily remove the maximum age restriction imposed in Section 2. The extremely old ages introduce additional computational cost to the estimation of the P-splines, potentially due to the limitations of the starting penalty in `bam`, and the training data were therefore reduced to 10 years for the sex-specific models.

Life expectancy at age 65

Model	Total	Male	Female
LC	20.63	18.81	22.15
LC_s	20.58	18.71	22.18
S_{tp}	20.43	19.08	21.43
S_{ps}	20.53	18.98	21.37
S_c	20.07	18.70	21.24

Table 13: Predicted life expectancy at age 65 in the year 2000

It is clear from Table 13 that life expectancy is higher for women than for men across all models. The Lee-Carter models estimate a greater spread between men and women, with differences of 3.3–3.4 years compared to 2.3–2.5 years for the spline models. All models predict very similar life expectancies for aggregated and male mortality, with the exception of the Currie P-spline, which predicts about half a year shorter aggregated life expectancy. The three spline models predict similarly for female life expectancy, with slightly lower predictions compared to the Lee-Carter models.
It is not immediately obvious which model estimates life expectancies with greater precision as enough empirical data do not exist. However, it is possible to limit the life expectancy calculations to 2000–2022 and compare them with the empirical life expectancies during this period to evaluate the accuracy of the predictions for the first 22 years of forecasting.

	Life expectancies			Relative error (%)		
Model	Total	Male	Female	Total	Male	Female
Empirical	18.08	17.16	18.95	–	–	–
LC	17.64	16.73	18.46	**-2.45**	-2.55	-2.58
LC_s	17.61	16.65	18.47	-2.61	-3.02	**-2.53**
S_{tp}	17.59	16.84	18.23	-2.74	**-1.91**	-3.78
S_{ps}	17.61	16.80	18.16	-2.62	-2.13	-4.17
S_c	17.43	16.64	18.14	-3.57	-3.03	-4.30

Table 14: Predicted life expectancy at age 65 in the year 2000, with calculations limited to years with empirical data

The results are displayed in Table 14 and indicate that all models underestimated the life expectancies at age 65 in the year 2000 and that the Currie P-spline provides the least accurate life expectancy predictions across all models and genders. This means that the models largely overestimated mortality in the years 2000–2022 for a person aged 65 in year 2000. For the remaining four models, the aggregated life expectancies are very similar, although the spline models produce more accurate predictions for males, while the Lee-Carter models perform better for females.

It is also possible to estimate the proportion of 65-year-olds in 2000 who survived the full 22 years and compare it with the observed proportion, with the survival rates shown in Table 15.

	Survival(%) at 2022			Relative error (%)		
Model	Total	Male	Female	Total	Male	Female
Empirical	45.74	38.48	52.63	–	–	–
LC	45.05	36.73	52.14	-1.51	-4.55	-0.94
LC_s	44.94	36.18	52.43	-1.74	-5.98	**-0.39**
S_{tp}	44.80	38.74	49.48	-2.06	0.66	-5.99
S_{ps}	45.25	38.26	48.34	**-1.07**	**-0.59**	-8.15
S_c	42.72	36.32	48.24	-6.59	-5.63	-8.35

Table 15: Predicted survival for a 65-year-old person by 2022

The Currie P-spline again demonstrates poor accuracy for survival forecasts for both aggregated and sex-specific models. It is also clear that, with the exception of the Currie P-spline, spline models estimate male survival more accurately, whereas the Lee-Carter models better predict female survival. For the aggregated models, all models, with the exception of the Currie P-spline, calculate similar and accurate survival estimates.

As with the life expectancy comparisons, the models estimate the survival rate to be lower than the observed survival rates. Plot 7 illustrates these mortality forecasts for a 65-year-old person in the year 2000 against empirical mortality. From the plot, it can be seen that the Currie P-spline largely overestimates mortality compared to

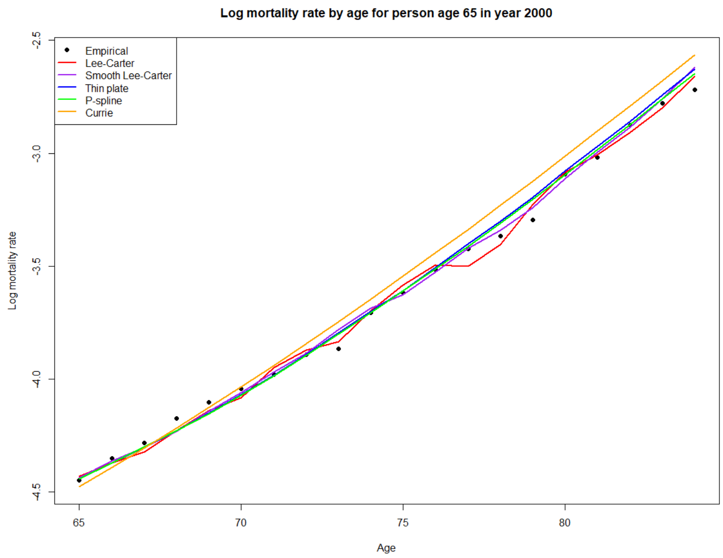


Figure 7: Log mortality for a person aged 65 in year 2000 for the next 22 years

the other models and the observed mortality rates, leading to substantially different life expectancy and survival predictions.

As a final note, it is important to remember that there are insufficient empirical data to calculate the true life expectancies for a person aged 65 in year 2000. Consequently, while the Currie P-spline provides the least accurate forecasts at ages 65–87, it may still produce the most accurate forecasts beyond 87.

4.4 Mortality Study

The primary objective has been to examine the forecasting accuracy of spline-based models for insurance purposes. However, evaluating forecasts requires observations beyond the training period, which leads to fitting models using older data. Consequently, there is a tradeoff between fitting models using older data, which allows for strong forecasting analysis, and using more modern data without future observations available for assessing prediction errors. As the original objective of examining forecasting accuracy has been completed, the study concludes with a mortality study of the models fitted to the full modern dataset up to the year 2022. Since there are no future observations past the year 2022, the statistical fit measures specified in Section 3.6 are evaluated against the training data instead of future observations, and insurance prices are reported in kronor rather than as error terms. The comparison focuses on comparing metrics and pricing for the different models, although it will not be possible to tell which model provides the most accurate forecasts. All models and training lengths are as specified in Sections 4.1 and 4.2.

4.4.1 Aggregated Mortality

	2013–2022		
Model	Deviance	RMSE	MAE
LC	**0.49**	**33.05**	**19.56**
LC_s	0.71	48.36	25.22
S_{tp}	0.74	47.42	25.46
S_{ps}	0.75	47.57	25.55
S_c	2.01	55.19	35.3

Table 16: Statistical fit for aggregated mortality on training data from 2013–2022

The statistical fit metrics in Table 16 clearly show that the standard Lee-Carter model provides the strongest fit on the training data, with the smoothed Lee-Carter model, thin plate spline, and P-spline providing very similar, although slightly worse, interpolation and fit. The Currie P-spline provides a significantly weaker statistical fit for all metrics.

To improve readability, the prices for the 20-year term life insurance contracts have been scaled up by a factor of 100 and can be seen in Table 17.

	Term life price			
Model	40	50	60	70
LC	2.17	6.49	17.78	44.28
LC_s	2.22	6.55	17.85	44.78
S_{tp}	2.23	6.55	17.75	42.89
S_{ps}	2.31	6.61	17.33	41.99
S_c	2.40	6.21	17.46	45.68

Table 17: Pricing for aggregated 20-year term life insurance policies by issue ages using 10 years of training data

The first thing to note is the increase in pricing with age. This is expected, as term life insurance pays out on death, which is more common at older ages. Next, it should be noted that the models do not agree regarding pricing, as the prices vary by 3–10% between the lowest and highest estimates, and that the standard Lee-Carter model predicts lower term life insurance prices for younger ages where these products are most common. The greatest variation in pricing estimates occurs at age 40, where the Currie P-spline produces the highest prices, approximately 10% above those of the standard Lee-Carter model, which produces the lowest estimates. At age 50, the pricing estimates are closer, with only the Currie P-spline producing slightly lower estimates, resulting in a 6.4% spread between the lowest and highest estimates. At age 60, the estimates are even more similar, with all models producing prices within 3% of one another, although the two P-splines still produce noticeably lower estimates than the remaining models. Age 70 produces larger differences, with the thin plate and P-spline estimating lower prices and the Currie P-spline estimating higher prices, resulting in a 6.5% difference between the models.

As life annuities are much larger in scale, life annuity prices have been reported without any rescaling and can be found in Table 18.

	Life annuity price			
Model	40	50	60	70
LC	16.20	15.94	15.17	13.25
LC_s	16.20	15.94	15.16	13.23
S_{tp}	16.20	15.93	15.17	13.27
S_{ps}	16.20	15.92	15.18	13.33
S_c	16.18	15.94	15.22	13.19

Table 18: Pricing for aggregated 20-year life annuities across issue ages using 10 years of training data

It is immediately clear that the models agree on pricing for life annuities to a greater degree than for term life insurance. This is expected, as term life pricing is more sensitive to mortality predictions than life annuities. At ages 40, 50, and 60, all models largely produce similar pricing, with the Currie P-spline differing by 0.1–0.3%. The biggest disagreement occurs at age 70, with the P-spline estimating slightly higher insurance pricing and the Currie P-spline estimating slightly lower, with around 1% pricing differences across the models. The Lee-Carter models and the thin plate spline, to a large extent, agree on annuity pricing across all ages.

4.4.2 Sex-Specific Mortality

As previous tests have shown, sex-specific mortality models require longer training periods than the aggregated models, and as such, the sex-specific mortality models use 20 years of training data from 2003–2022. The statistical fit for the models is shown in Table 19.

	Male			Female		
Model	Deviance	RMSE	MAE	Deviance	RMSE	MAE
LC	**1.02**	**23.13**	**14.6**	**1.01**	**22.14**	**13.05**
LC_s	1.28	29.21	17.25	1.14	25.88	14.54
S_{tp}	1.15	26.88	16.08	1.12	25.13	14.26
S_{ps}	1.19	27.27	16.34	1.16	25.48	14.5
S_c	1.21	27.25	16.32	1.16	25.48	14.5

Table 19: Statistical fit for female mortality on training data from 2003–2022

One notable observation is the better statistical fit for the female mortality models compared to the male models, particularly for the MAE, suggesting that the female

mortality models fit the training data better. This does not necessarily mean better and more accurate forecasts, which is the primary objective for mortality models for insurance purposes, but could indicate lower variance in the training data for women than for men. The gender-separated mortality models also identify the standard Lee-Carter model as providing the strongest statistical fit across all metrics and genders. The Currie P-spline produces better fits for the sex-specific models than for the aggregated models, as the statistical fit is closer to that of the other models and no longer deviates from the other two spline models. The smoothed Lee-Carter model provides a slightly worse statistical fit for the male mortality models.
For the female mortality models, the only deviation is a stronger statistical fit for the standard Lee-Carter model compared to the other models. It can also be seen that the thin plate spline is the best-performing spline model in terms of statistical fit, as it fits the data better for all metrics and genders.
The statistical fit of the sex-specific models differs from that of the aggregated models shown in Table 16, with the aggregated models exhibiting lower Poisson deviance, but higher RMSE and MAE values. As these metrics evaluate the difference between observed and predicted death counts, the lower Poisson deviance for the aggregated models is unexpected, since the aggregated death counts are, on average, twice as large. One would therefore expect larger discrepancies and, consequently, larger fit metrics for the aggregated models. This difference is explained by the models being estimated on different datasets, as the aggregated models use 10 years of training data whereas the sex-specific models use 20 years. Since the fit metrics are evaluated on the training data, differences in the datasets also affect the resulting metrics. To allow for a proper comparison, the aggregated models were re-estimated using the same training period as the sex-specific models, with the updated fit metrics reported in Table 20.

	2003–2022		
Model	Deviance	RMSE	MAE
LC	**1.1**	**35.69**	**21.15**
LC_s	1.4	45.77	25.04
S_{tp}	1.28	42.92	23.4
S_{ps}	1.32	43.39	23.85
S_c	1.33	43.28	23.82

Table 20: Statistical fit for aggregated mortality on training data from 2003–2022

These results align more closely with expectations, with the sex-specific models exhibiting lower deviance and better statistical fit than the aggregated models.

As in the aggregated mortality study, all term prices have been scaled up by a factor of 100 to improve readability and are shown in Table 21.

Model	Male 40	Male 50	Male 60	Male 70	Female 40	Female 50	Female 60	Female 70
LC	2.49	7.25	18.57	47.99	1.71	5.15	14.70	39.86
LC_s	2.47	7.15	18.39	47.74	1.68	5.09	14.58	39.66
S_{tp}	2.67	8.04	21.65	50.12	1.67	5.10	15.23	40.24
S_{ps}	2.53	8.13	22.16	49.80	1.64	5.04	15.33	40.26
S_c	2.54	8.11	21.97	49.71	1.64	5.04	15.32	40.24

Table 21: Pricing for 20-year term life insurance policies by issue ages

One interesting observation is the lower prices for women compared to men, which is expected considering the lower mortality rates of women. In general, the female mortality models produce less spread in price estimates, only differing by 1–5% between the lowest and highest pricing estimates without any clear outliers. Compared with the other spline models, the thin plate spline more closely resembles the Lee-Carter models, estimating prices between those of the two P-splines and the Lee-Carter models for all ages. The spline models, in general, estimate lower prices for younger ages and higher prices for older ages compared to the Lee-Carter models, with the two P-splines estimating very similar prices. It can also be seen that the female Lee-Carter models estimate very closely, with the smoothed Lee-Carter model estimating slightly lower than the standard one.
For the male mortality models, there are considerably larger differences in price estimates between the models, with differences of 5–20% between the lowest and highest estimates. The Lee-Carter models estimate lower prices across all ages, with the smoothed Lee-Carter model estimating the lowest. Both P-splines estimate similar prices, although higher than those of the Lee-Carter models, with the Currie P-spline producing the lower pricing estimates of the two. The thin plate spline pricing estimates deviate from the other two spline models, providing higher estimates at ages 40 and 70, but lower estimates at ages 50 and 60.
Across both genders, it is clear that the spline models estimate higher term prices compared to the Lee-Carter models, but as there are no future observations, it is not possible to know if this is an overestimation of prices or if the higher prices are appropriate.

Finally, the 20-year life annuity prices are calculated using the models and are shown in Table 22.

Model	Male 40	Male 50	Male 60	Male 70	Female 40	Female 50	Female 60	Female 70
LC	16.18	15.87	15.05	12.98	16.24	16.02	15.39	13.68
LC_s	16.18	15.88	15.07	13.00	16.24	16.03	15.40	13.70
S_{tp}	16.17	15.86	14.88	12.72	16.24	16.03	15.39	13.63
S_{ps}	16.18	15.86	14.87	12.72	16.24	16.03	15.38	13.62
S_c	16.18	15.86	14.87	12.72	16.24	16.03	15.39	13.62

Table 22: Pricing for 20-year life annuities across issue ages

Here it is possible to see that the life annuity prices for women are higher across all ages, which is to be expected considering that life annuities pay out on survival and women have lower mortality rates and longer life expectancies. As with the aggregated models, the price differences between models are very low, with prices differing by at most 2% for men and 0.5% for women. As with term prices, the price ranges are closer for the female mortality models than for the male models, which is expected, as the statistical fit metrics potentially suggest higher variance in male mortality during the training period. As with the term life insurance prices, the spline models estimate higher mortality than the Lee-Carter model, as can be seen in the lower annuity prices.

4.4.3 Life Expectancy Estimates

As with the pre-2000 models discussed in Section 4.2 and 4.3, it is possible to calculate the life expectancy of a 65-year-old, although it is necessary to remove the maximum age restriction. The P-splines no longer prove to be infeasible for the preferred 20-year training period for the sex-specific models, and are therefore no longer restricted to only 10 years of data as was done with the pre-2000 life expectancy models. The resulting life expectancies at age 65 are shown in Table 23.

Life expectancy at age 65

Model	Total	Male	Female
LC	22.77	21.80	23.90
LC_s	22.71	21.88	23.96
S_{tp}	22.97	20.99	23.88
S_{ps}	22.56	21.00	23.90
S_c	22.55	21.01	23.89

Table 23: Predicted life expectancy at age 65 in the year 2022

Compared to the pre-2000 models, all models project higher life expectancies across both aggregated and sex-specific data, with increases of 2–2.5 years for the aggregated models, 1.9–3.2 years for males, and 1.7–2.6 years for females. This suggests that mortality rates at older ages have declined over the last 22 years. The models still project higher life expectancies for women compared to men, although the Lee-Carter models estimate a smaller gender gap than in 2000, at 2.1 years compared to 2.3–2.5 years. All models produce similar life expectancy estimates for women, but the spline models estimate much lower life expectancy for men compared to the Lee-Carter models, resulting in gender gaps of around 2.9 years for the spline models compared to 2.3–2.5 years in 2000. Figure 8 highlights this issue, where it is possible to see that the years leading up to 2022 included high variance in mortality for men aged 65, potentially related to the COVID-19 pandemic. Spline models are not designed for extrapolation and the extrapolation behavior is a side effect of how the models behave outside of their boundary, with the model stretching out the surface in an unpenalized direction. As the models are penalized by second-order derivatives, mortality is extrapolated linearly. As a consequence, large variances

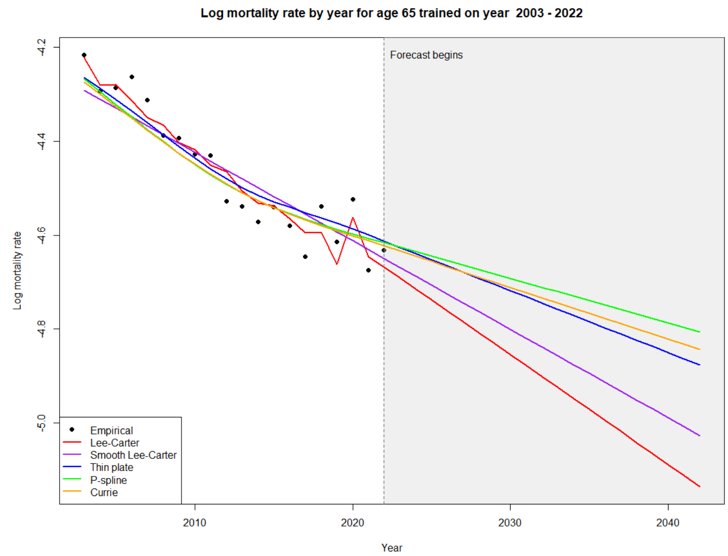


Figure 8: Log mortality forecasts for males aged 65 based on 20 years of training data

near the edges of the training data can result in the model reaching the boundary in an undesirable direction and cause poor extrapolation.

The aggregated life expectancy is largely consistent across models, with both Lee-Carter models producing estimates within ±0.2 years of those from the spline models. A notable result is that the smoothed Lee-Carter model estimates that male and female life expectancies are longer than the standard Lee-Carter model while still estimating lower aggregated life expectancy.

By extrapolating the mortality rates from the pre-2000 models used in the life expectancy analysis in Section 4.3.4, it is possible for the pre-2000 models to estimate the life expectancy of a 65-year-old in 2022 and compare it with the 2022 models. Table 24 shows these life expectancies and their relative differences compared to those calculated using 22 years of additional data.

	2022			Extrapolations			Relative error (%)		
Model	Total	Male	Female	Total	Male	Female	Total	Male	Female
LC	22.77	21.80	23.90	23.08	21.51	24.20	1.36	-1.27	1.28
LC_s	22.71	21.88	23.96	22.97	21.35	24.22	1.15	-2.42	1.07
S_{tp}	22.97	20.99	23.88	22.68	22.30	22.48	-1.26	6.25	-5.85
S_{ps}	22.56	21.00	23.90	22.93	22.08	22.15	1.61	5.12	-7.3
S_c	22.55	21.01	23.89	21.88	21.34	22.29	-3.00	1.55	-6.72

Table 24: Calculated life expectancies at age 65 compared against forecasted life expectancies

From this, it is clear that the life expectancy extrapolations from the pre-2000 models

are relatively consistent with those obtained using the extended dataset, with only the Currie P-spline differing by more than 2%. However, larger differences are observed for male and female life expectancies, particularly for the spline models, with the Currie P-spline being the only model whose estimates change by less than 2%. The Lee-Carter models show greater stability than the spline models, with life expectancy estimates differing by only 1–2%.
It is worth noting that this does not necessarily prove that the Lee-Carter models provide more accurate mortality projections, only that they provide greater stability in forecasts, particularly for the models separated by gender.

4.5 Mortality Projection

The analysis concludes with several 50-year mortality forecasts for ages 30, 40, 50, 60, 70, and 80, which can be found in Figure 9. The plots are separated into two for improved readability, although, if preferred, the combined plot can be found in Appendix C.

4.5.1 Aggregated Forecasts

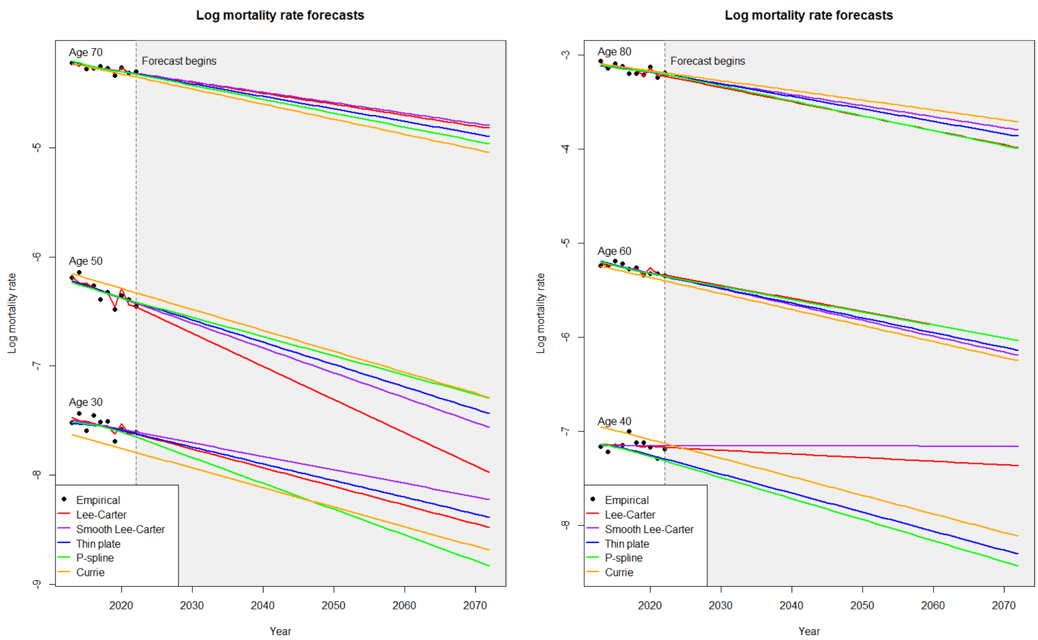


Figure 9: 50-year forecasts for aggregated mortality

Plot 9 shows a clear downward trend for all models and ages, with the models forecasting 0.5–1 lower log mortality over the next 50 years, resulting in 26–63% lower mortality rates. The exception is the Lee-Carter models at age 40, which project, at most, a very small decrease in mortality, and even project that mortality rates at age 40 will be higher than at age 50 by the year 2055. Several other mortality forecasts intersect across models and can be seen in Figure 12 in Appendix C. Most notable are the thin plate and P-spline at age 40, which project lower mortality than the smoothed Lee-Carter model and mortality very close to the thin plate spline at age 30 over the next 50 years. It can also be seen that the difference in mortality forecasts at older ages, particularly at age 60, is smaller, which is to be expected, as mortality has been plotted on the log scale, which compresses differences at higher mortality rates.

4.5.2 Male Forecasts

As with the aggregated models, the forecasts have been separated to improve readability with the combined plot in Appendix C. It is immediately clear from Figure

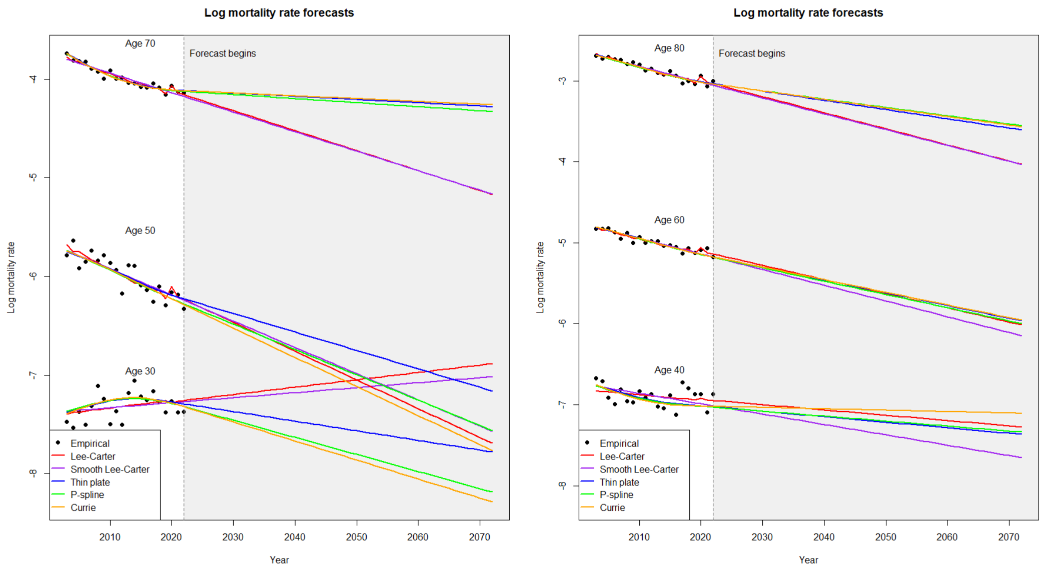


Figure 10: 50-year forecasts for male mortality

10 that the male mortality models do not agree on future mortality, and this highlights the problem with spline extrapolation: it is unpredictable. One of the clearest examples of this is at age 70, where the observed log mortality shows a clear linear downward trend, but the spline models still project a near-constant mortality rate due to the increased mortality near the boundary. The mortality shock observed during the COVID-19 period can be seen across the older ages, with the spline models projecting higher mortality than the more stable Lee-Carter projections. This sensitivity near the boundary can also be seen at age 30, where the Lee-Carter models project an increase in mortality, but the spline models estimate a decrease, as the mortality near the boundary was on a downward trend.

It is not clear which model provides the most accurate forecasts, but the spline models' sensitivity to variance near the boundaries is concerning. As with the aggregated models, the Lee-Carter models expect male mortality at younger ages to overtake mortality at older ages, although to a much greater extent, expecting mortality at age 30 to overtake mortality at age 40 by 2050 and mortality at age 50 by 2055. Furthermore, the standard Lee-Carter model and the P-splines project that mortality at age 40 will overtake the mortality at age 50 by 2060, but the smoothed Lee-Carter model and the thin plate spline do not.

4.5.3 Female Forecasts

The forecasts for the female models are also separated into two for readability, with the full plot attached in Appendix C. The plots in Figure 11 show that the

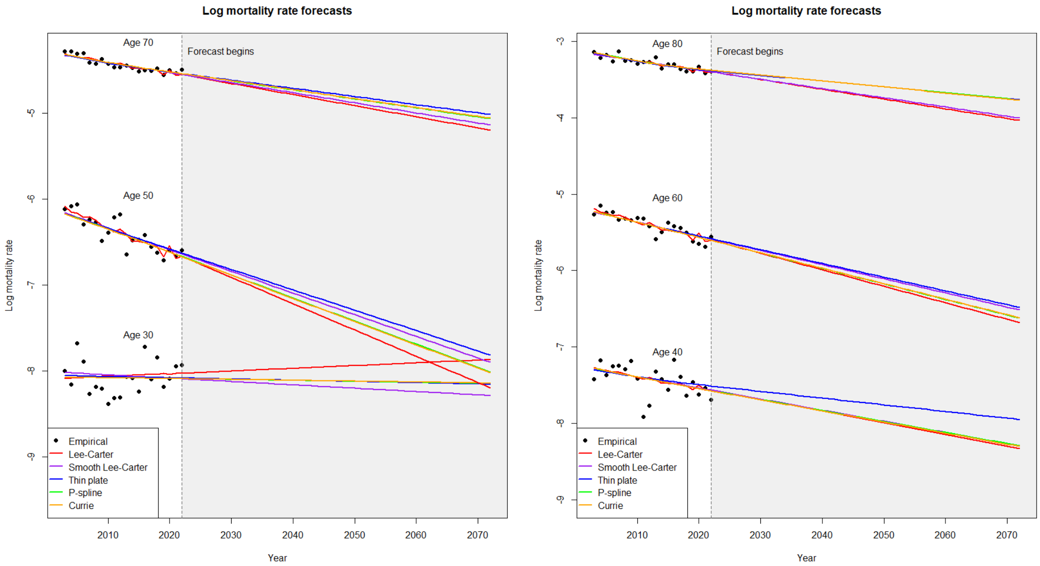


Figure 11: 50-year forecasts for female mortality

models have more similar forecasts, with only a few notable exceptions. The first is the relatively spread-out forecasts at age 50, and the thin plate spline projecting notably higher mortality at age 40. There is also a noteworthy difference between the projections of the standard and the smoothed Lee-Carter models, as the two models have provided very similar statistical fits, pricing, and forecasts throughout this study. At age 30, however, the smoothed Lee-Carter model projects a slight decrease in mortality, while the standard Lee-Carter model expects an increase in mortality. As with the aggregated and male models, the standard Lee-Carter model predicts the mortality at younger ages to overtake that of older ages. In this case, mortality at age 30 is expected to overtake mortality at age 40 by 2050 and mortality at age 50 by 2065. The P-splines also project that mortality at age 30 will overtake mortality at age 40 by 2060, but otherwise no models or age groups are expected to overtake those of older ages However, the thin plate spline projects that mortality at ages 30, 40, and 50 will become very similar over the next 50 years.

5 Results

The analysis so far has focused on comparing the different models across different tests, periods, and metrics, although it might not be immediately clear what this implies for the different models. This section will focus on the individual models and discuss what the conducted tests reveal about their performance.

5.1 Lee-Carter

Both Lee-Carter models have proven to be very robust, and it is clear why the Lee-Carter model is widely used in mortality modeling. The standard Lee-Carter model handled the full dataset well, only struggling with older ages, but the smoothed model did not, as the linear year trend assumption was not satisfied. The Lee-Carter models struggled with the shorter training lengths for the sex-specific data, although this issue affected all models.

The pricing was accurate for both models, and the Lee-Carter models provided more accurate pricing estimates than the spline models for most products and ages for the aggregated dataset, with the standard Lee-Carter model performing better at younger ages and the smoothed model performing better at older ages. The Lee-Carter models struggled with the gender-separated models, providing poor forecasts for males at older ages and females at younger ages. With the male data, the models' age preference changed compared to the aggregated models, with the smoothed Lee-Carter model estimating younger ages better and the standard Lee-Carter model estimating older ages better. For women, the smoothed Lee-Carter model was preferred across all ranges except for age 70.

For the models fitted with the full modern data, the standard Lee-Carter model provided better statistical fit for both aggregated and sex-specific models. Compared to the smoothed Lee-Carter model, the standard Lee-Carter model predicted lower mortality for the aggregated data, leading to lower term life insurance prices, higher life annuity prices, and longer life expectancy. For the gender-specific models, the standard Lee-Carter model predicted higher mortality for both genders, resulting in higher term life insurance prices, lower annuity prices, and lower life expectancies.

The 50-year mortality forecasts showed no major differences between the models, with the only noteworthy distinction being the standard model projecting lower mortality for the full population and higher mortality when genders were separated, as well as the disagreement regarding whether mortality for 30-year-old women was projected to increase or decrease.

5.2 Splines

Despite the lack of theoretical justifications for spline extrapolation, the spline models performed surprisingly well. The spline models provided better statistical fit and pricing predictions than the Lee-Carter models for longer training periods, while also producing accurate estimates for shorter training periods for the aggregated dataset, although these were slightly less accurate than those of the Lee-Carter models. As with the Lee-Carter models, the spline models struggled with the shorter training periods for the sex-specific data and required at least 20 years of data to estimate prices accurately.

It is not clear which spline model produced the most accurate insurance prices for the aggregated models, as the P-spline provided the best estimates for younger ages, while the thin plate performed best for older ages. For the male models, the splines provided more accurate pricing estimates at older ages relative to the Lee-Carter models, and provided similar, although slightly less accurate, pricing for younger ages. It is, however, not clear which spline model is preferred for pricing male insurance products, as the most accurate model depended on the age of the insured. The pricing estimates of the standard P-spline were very accurate for women, either being the most accurate of all models, including the Lee-Carter models, or providing estimates close to those of the most accurate models. In general, the spline models provided more accurate overall pricing than the Lee-Carter models for the gender-separated data. The thin plate and P-spline provided accurate pricing estimates across all ages and datasets and either provided better or comparable pricing estimates relative to the Lee-Carter models. The Currie P-spline also estimated reasonably accurate insurance prices, although slightly less accurately than the other two spline models.

The thin plate and P-spline calculated the life expectancy of a 65-year-old in 2000 very similarly to the Lee-Carter models for the aggregated dataset, while the Lee-Carter models provided better life expectancy estimates for females and the thin plate and P-spline provided more accurate estimates for males. The Currie P-spline was less accurate than all other models in its life expectancy estimates.

For the models fitted with the full modern data, the thin plate and P-spline provided a similar fit to the smoothed Lee-Carter model, but the Currie P-spline model fit was significantly worse for the aggregated data. The pricing for the thin plate spline was very similar to that of the Lee-Carter models for most ages, but the spline estimated lower mortality for older ages, resulting in lower term life insurance prices and higher annuity prices. This effect was also seen in the P-spline to a greater extent, providing even lower term prices and higher annuity prices for older ages. The Currie P-spline estimated higher mortality for most ages, resulting in higher term prices and lower annuity prices.
The statistical fit of the Currie P-spline was better for the sex-specific models than for the aggregated models, resulting in statistical fit metrics very similar to those of the other spline models and the smoothed Lee-Carter model. The spline models predicted higher mortality than the Lee-Carter models for males across all ages, resulting in higher term prices and lower annuity prices. For the female models, the spline models estimated slightly higher mortality at older ages compared to the Lee-Carter models, and slightly lower mortality for younger ages. Life expectancy estimates for the aggregated data were slightly higher for the thin plate spline than for the Lee-Carter models, while the P-splines produced slightly lower estimates. Female life expectancies were estimated to be very close to those of the Lee-Carter models, but male life expectancies were estimated to be significantly lower.

The 50-year mortality projections show no clear pattern as to which spline model provides the best forecasts, and it depends entirely on age and gender which spline estimates lower or higher mortality rates. For the aggregated dataset, the forecasts align decently well with the Lee-Carter models. However, the reduced exposure in the sex-specific data introduces additional variance, leading to less predictable forecasts for the sex-specific models, as the models attempt to fit the increased variance and the mortality shock observed during the COVID-19 period. This further highlights the limited theoretical foundation for extrapolation of spline-based models.

It is not immediately clear whether the stable Lee-Carter models or the flexible spline models produced the overall most accurate forecasts, although it is clear that the spline-based models have a lot of potential and often outperformed the Lee-Carter models in pricing accuracy. It is unfortunately also clear that the unpredictability of spline extrapolation can make the spline-based models very risky, particularly for price-setting purposes. While it is possible to mitigate these risks, it highlights the benefits of stable mortality models such as the Lee-Carter model and helps explain its widespread use.

6 Discussion

The objective of this study was to examine the Lee-Carter model and compare potential alternatives to evaluate how accurately more flexible models can price insurance products. This evaluation was performed using out-of-time validation with varying training lengths and primarily focused on 20-year term life insurance and 20-year life annuity pricing errors, although a brief analysis was also conducted for statistical fit and mortality forecasts for modern data.

The results have demonstrated why the Lee-Carter model, together with its smoothed counterpart, is considered a standard mortality model, as both have been shown to be more stable for pricing insurance products across all ages, and at worst, provided only slightly less accurate pricing for sex-specific models. For statistical fit, the differences are smaller with no clear favorite. The spline models still provided very strong pricing accuracy, particularly for the aggregated data, and only struggled with older ages and the reduced exposure in the sex-specific data. However, the models fit with the full modern data also illustrated the unpredictability of spline extrapolation, as the increased mortality observed during the COVID-19 period introduced additional shock near the boundaries.

The differences in pricing errors are small enough that further research in spline forecasting could potentially improve the spline models and challenge the Lee-Carter models for actuarial applications. One such area that has the potential to yield interesting results is further experimentation with the number of knots and knot locations, and with stronger penalties for wiggliness to force more linear time trends. The structure of the tensor product spline with a smooth function for age and a tensor product for age and year, such as the spline models proposed in this study, is similar to the structure of the Lee-Carter model. Further studies in this area could lead to some interesting theoretical extrapolation results, potentially by smoothing age while treating year as a time series to preserve the flexibility of spline models while gaining a better theoretical foundation for extrapolation. Similarly, it is possible that smoothing the Lee-Carter age parameter β_x with splines could yield some interesting results, as the smoothed Lee-Carter model produced more accurate price estimates for certain ages and genders.

References

Boor, C. de (1978). *A Practical Guide to Splines*. Vol. 27. Applied Mathematical Sciences. New York: Springer. DOI: `https://doi.org/10.2307/2006241`.

Booth, H. et al. (2006). "Lee–Carter mortality forecasting: a multi-country comparison of variants and extensions". In: *Demographic Research* 15.9, pp. 289–310. DOI: `10.4054/DemRes.2006.15.9`. URL: `https://www.demographic-research.org/volumes/vol15/9/`.

Continuous Mortality Investigation (2011). *User Guide to Version 1.3 of the CMI Library of Mortality Projections*. User Guide. Institute and Faculty of Actuaries. URL: `https://www.actuaries.org.uk/system/files/documents/pdf/cmi-projectionslibraryuserguidev1-3.pdf`.

Currie, I. D., Durban, M., and Eilers, P. H. C. (2004). "Smoothing and Forecasting Mortality Rates". In: *Statistical Modelling* 4.4, pp. 279–298. DOI: `10.1191/1471082X04st080oa`.

Duchon, J. (1977). "Splines Minimizing Rotation-Invariant Semi-Norms in Sobolev Spaces". In: *Constructive Theory of Functions of Several Variables*. Berlin: Springer, pp. 85–100. DOI: `10.1007/BFb0086566`.

Eilers, P. H. C. and Marx, B. D. (1996). "Flexible Smoothing with B-Splines and Penalties". In: *Statistical Science* 11.2, pp. 89–121. DOI: `10.1214/ss/1038425655`.

Hall, M. and Friel, N. (2011). "Mortality Projections using Generalized Additive Models with Applications to Annuity Values for the Irish Population". In: *Annals of Actuarial Science* 5.1, pp. 19–32. DOI: `10.1017/S1748499510000011`.

Hastie, T. and Tibshirani, R. (1986). "Generalized Additive Models". In: *Statistical Science* 1.3, pp. 297–310. DOI: `10.1214/ss/1177013604`.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Vol. 43. Monographs on Statistics and Applied Probability. Boca Raton, FL: Chapman and Hall/CRC. ISBN: 9780203753781. DOI: `10.1201/9780203753781`.

Human Mortality Database (2024). *Human Mortality Database*. University of California, Berkeley and Max Planck Institute for Demographic Research. URL: `https://www.mortality.org` (visited on 02/01/2024).

Lee, R. D. and Carter, L. R. (1992). "Modeling and Forecasting U.S. Mortality". In: *Journal of the American Statistical Association* 87.419, pp. 659–671. DOI: `10.1080/01621459.1992.10475265`.

Nelder, J. A. and Wedderburn, R. W. M. (1972). "Generalized Linear Models". In: *Journal of the Royal Statistical Society: Series A* 135.3, pp. 370–384. DOI: `10.2307/2344614`.

Patterson, H. D. and Thompson, R. (1971). "Recovery of Inter-Block Information when Block Sizes are Unequal". In: *Biometrika* 58.3, pp. 545–554. DOI: `10.1093/biomet/58.3.545`.

Reiss, P. T. and Ogden, R. T. (2009). "Smoothing Parameter Selection for a Class of Semiparametric Linear Models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, pp. 505–523. DOI: `10.1111/j.1467-9868.2008.00695.x`.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press. DOI: `10.1017/CBO9780511755453`.

Svensk Försäkring (2021). *Försäkrade i Sverige: Livslängder och dödlighet – prognoser 2021–2080 (DUS21)*. Svensk Försäkring. URL: https://www.svenskforsakring.se/globalassets/publikationer-arkiv/dus-21/forsakrade-i-sverige-livslangder-och-dodlighet-prognoser-2021-2080-dus21.pdf/.

Svensk Försäkring (2023). *Dödlighetsundersökningen 2023 (DUS23): En vidareutveckling av DUS21*. Svensk Försäkring. URL: https://www.svenskforsakring.se/globalassets/publikationer/dus-23/dodlighetsundersokningen-2023-dus23---en-vidareutveckling-av-dus21-reviderad-2025-04-08.pdf/.

Villegas, A. M., Kaishev, V. K., and Millossovich, P. (2023). *StMoMo: An R Package for Stochastic Mortality Modeling*. CRAN. URL: https://CRAN.R-project.org/package=StMoMo.

Wood, S. N. (2003). "Thin Plate Regression Splines". In: *Journal of the Royal Statistical Society: Series B* 65.1, pp. 95–114. DOI: 10.1111/1467-9868.00374.

Wood, S. N. (2006). "Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models". In: *Biometrics* 62.4, pp. 1025–1036. DOI: 10.1111/j.1541-0420.2006.00574.x.

Wood, S. N. (2011). "Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.1, pp. 3–36. DOI: 10.1111/j.1467-9868.2010.00749.x.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*. 2nd ed. Texts in Statistical Science. Boca Raton: Chapman and Hall/CRC. DOI: 10.1201/9781315370279.

Wood, S. N. (2023). *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*. R package version. URL: https://CRAN.R-project.org/package=mgcv.

Wood, S. N., Goude, Y., and Shaw, S. (2015). "Generalized Additive Models for Large Data Sets". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 64.1, pp. 139–155. DOI: 10.1111/rssc.12068.

A Model Implementation in R

Model fitting assumes a predefined dataset called `train_data` which includes `deaths`, `exp`, `Age`, and `Year`. It is also necessary to define start and end points for the training and testing years with `train_start`, `train_end`, `test_start`, and `test_end`. The spline models are then fitted with `GAM` from the `MGCV` package in `R`.

```
library(mgcv)

ageRange = max(train_data$Age)-min(train_data$Age)
yearRange = train_end-train_start
kAge = round(ageRange/4)
kYear = min(round(yearRange/4),
            20)

spline_fit_tp <- bam(
 deaths ~
  s(Age,k=80) +
  te(Age,Year, bs = "tp", k = c(20,kYear)) +
  offset(log(exp)),
 data = train_data,
 nthreads = 8,
 family = quasipoisson(),
 discrete = TRUE,
 method = "fREML"
)

spline_fit_ps <- bam(
 deaths ~
  s(Age, bs = "ps", k = kAge) +
  te(Age,Year, bs = "ps", k = c(kAge,kYear), m=c(3,2)) +
  offset(log(exp)),
 data = train_data,
 nthreads = 8,
 family = quasipoisson(),
 discrete = TRUE,
 method = "fREML"
)

spline_fit_currie <- bam(
 deaths ~
  te(Age,Year, bs = "ps", k = c(kAge,kYear), m=c(3,2)) +
  offset(log(exp)),
 data = train_data,
 min.sp = c(1e-4,1e-4),
 nthreads = 8,
 family = quasipoisson(),
```

```
 method = "REML",
)
```

The standard Lee-Carter model is fitted with `fit` from the `StMoMo` package.

```
library(StMoMo)

Dxt <- with(train_data,
            tapply(deaths,
                   list(Age, Year),
                   sum))
Ext <- with(train_data,
            tapply(exp,
                   list(Age, Year),
                   sum))
ages  <- as.numeric(rownames(Dxt))
years <- as.numeric(colnames(Dxt))

lc_fit <- fit(
 lc(),
 Dxt = Dxt,
 Ext = Ext,
 ages = ages,
 years = years
)
```

The smoothed Lee-Carter model is then estimated by extracting the b_x and k_t parameters of the standard Lee-Carter model and smoothing them. Do note that the `rollmean` function from the `zoo` package is used to calculate the moving averages of the b_x parameter.

```
library(zoo)

bx_adj <- rollmean(lc_fit$bx, k = 5, fill = "extend")
years <- as.numeric(colnames(lc_fit$kt))
kt_train <- as.numeric(lc_fit$kt)
year_center <- mean(years)
kt_lm <- lm(kt_train ~ I(years - year_center))
```

The surface is then reconstructed with the new smoothed parameters.

```
fc_years <- test_start:test_end
years_all <- c(years, fc_years)

kt_all <- as.numeric(
 predict(
  kt_lm,
```

```
  newdata = data.frame(years = years_all)
 )
)

bx <- as.numeric(bx_adj)
ax <- as.numeric(lc_fit$ax)
kt <- as.numeric(kt_all)
lc_adj_surface <- exp(
 sweep(outer(bx, kt), 1, ax, "+")
)
dimnames(lc_adj_surface) <- list(
 Age  = ages,
 Year = years_all
)
```

Model extrapolation depends on the specific model, with the spline models extrapolating with `predict` and returning the number of deaths. As the spline models use an offset of `log(exp)`, the exposure must be strictly greater than 0. Letting the exposure be equal to 1 results in extrapolation returning mortality rates and is recommended unless exposure is known. In the case of predictions being given as numbers of deaths, mortality rates can be calculated by dividing by the exposure. For the standard Lee-Carter model, extrapolation is estimated with `fitted` and returns mortality rates. Extrapolation of the smoothed Lee-Carter model is done by examining the constructed surface, `lc_adj_surface`.

B Tables for Period Testing

B.1 Aggregated Mortality

1951–2000

Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	7.83	**92.03**	62.48	21.95	-11.01
LC_s	15.5	145.78	99.65	28.45	-17.03
S_{tp}	**6.31**	107.16	**52.85**	**2.39**	-2.96
S_{ps}	9.53	146.53	71.82	6.82	**-0.92**
S_c	7.77	115.01	54.74	2.85	-2.22

Table 25: Forecasting performance for aggregated mortality during 2001–2020 with 50 years of training data

1971–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	**4.09**	**69.98**	**44.42**	9.88	-7.84
LC_s	4.76	79.9	50.86	11.11	-9.12
S_{tp}	7.39	118.4	57.63	3.23	-2.93
S_{ps}	15.16	202.69	93.42	3.91	-3.25
S_c	10.8	166.32	66.72	**1.04**	**-2.24**

Table 26: Forecasting performance for aggregated mortality during 2001–2020 with 30 years of training data

1981–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	**3.32**	**56.68**	**34.44**	**2.72**	-4.08
LC_s	3.37	61.57	36.69	4.3	-4.99
S_{tp}	10.43	149.41	71.25	4.15	-3.38
S_{ps}	18.45	248.45	105.44	6.2	-5.07
S_c	13.27	181.6	79.48	3.75	**-3.06**

Table 27: Forecasting performance for aggregated mortality during 2001–2020 with 20 years of training data

B.2 Male Mortality

1751–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	29.98	206.14	112.6	5.74	-30.71
LC_s	133.43	422.73	282.24	79.72	-78.96
S_{tp}	**3.52**	**49.1**	**29.55**	**-4.04**	-7.72
S_{ps}	5.77	64.29	36.44	-4.26	**-6.81**
S_c	6.36	69.59	38.24	-6.41	**-6.81**

Table 28: Forecasting performance for male mortality during 2001–2020 with 250 years of training data

1951–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	11.82	113.41	64.89	16.24	-21.76
LC_s	39.18	218.08	136.55	39.4	-43.64
S_{tp}	5.58	62.81	36.53	**-5.06**	-6.1
S_{ps}	**5.01**	**52.24**	**31.7**	-6.26	**-0.89**
S_c	6.29	65.44	38.02	-5.54	-5.69

Table 29: Forecasting performance for male mortality during 2001–2020 with 50 years of training data

1971–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	6.98	87.63	48.68	3.89	-15.57
LC_s	10.34	110.14	64.32	10.14	-22
S_{tp}	6.42	71.01	40.08	**-3.47**	-6.81
S_{ps}	**6.13**	**63.15**	**38.04**	-5.92	**-4.22**
S_c	7.01	75.83	40.04	-6.41	-5.73

Table 30: Forecasting performance for male mortality during 2001–2020 with 30 years of training data

B.3 Female Mortality

1751–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	15.65	150.55	76.71	**-12.62**	-1.85
LC_s	133.96	447.58	275.08	76.31	-51.43
S_{tp}	**2.76**	**33.87**	**21.1**	20.23	**-1.22**
S_{ps}	3.7	44.47	25.52	18.46	1.77
S_c	3.7	45.04	25.52	19.05	1.85

Table 31: Forecasting performance for female mortality during 2001–2020 with 250 years of training data

1951–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	**2.59**	**37.38**	**21.76**	12.43	2.76
LC_s	3.48	53.44	28.08	**9.9**	4.71
S_{tp}	3.35	52.28	27.64	14.52	-0.26
S_{ps}	6.83	92.5	43.1	15.65	-0.73
S_c	5.95	81.36	39.11	15.48	**-0.19**

Table 32: Forecasting performance for female mortality during 2001–2020 with 50 years of training data

1991–2000					
Model	Deviance	RMSE	MAE	Term Error (%)	Annuity Error (‰)
LC	4.69	45.53	26.55	14.22	**-0.06**
LC_s	**2.93**	**36.06**	**21.28**	11.29	1.1
S_{tp}	7.21	88.92	42.96	27.74	-3.66
S_{ps}	7.09	94.88	47.02	17.22	-6.07
S_c	5.73	80.44	41.87	**6.3**	-6.05

Table 33: Forecasting performance for female mortality during 2001–2020 with 10 years of training data

C Mortality Forecasting Plots

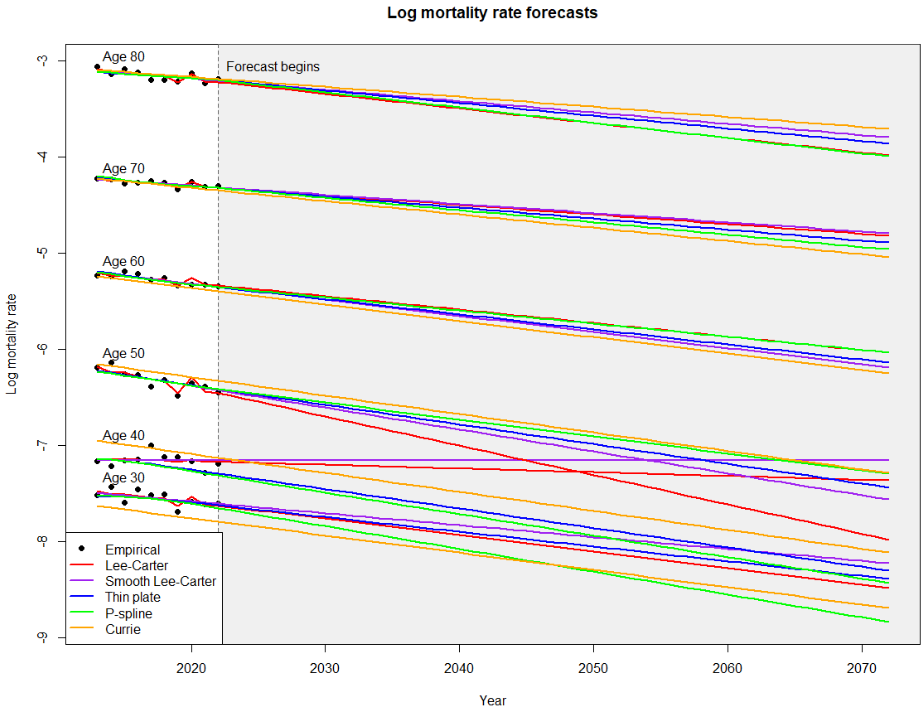


Figure 12: 50-year forecasts for aggregated mortality

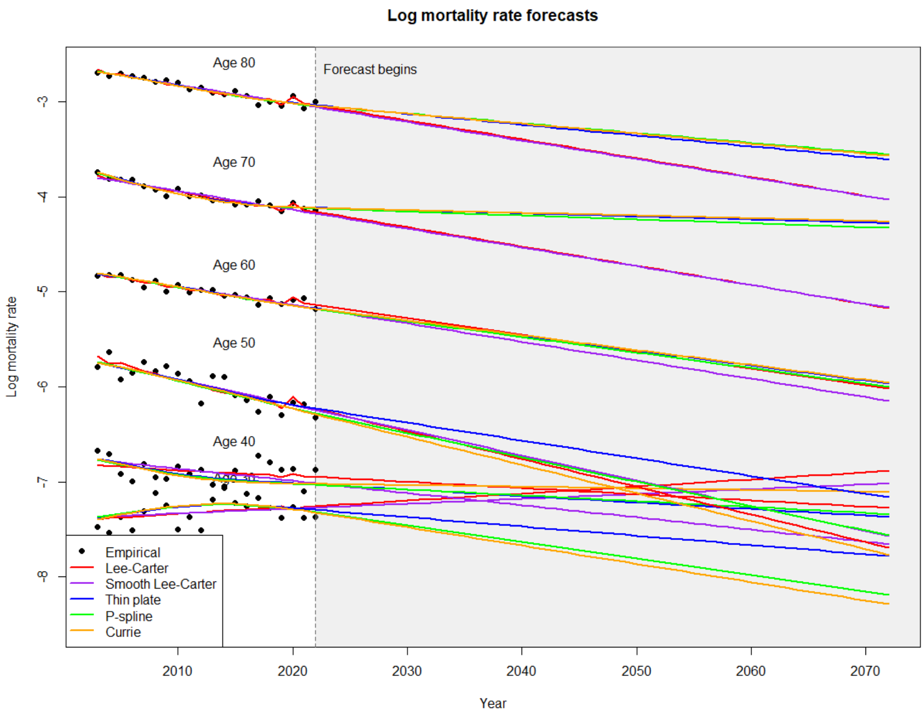


Figure 13: 50-year forecasts for male mortality

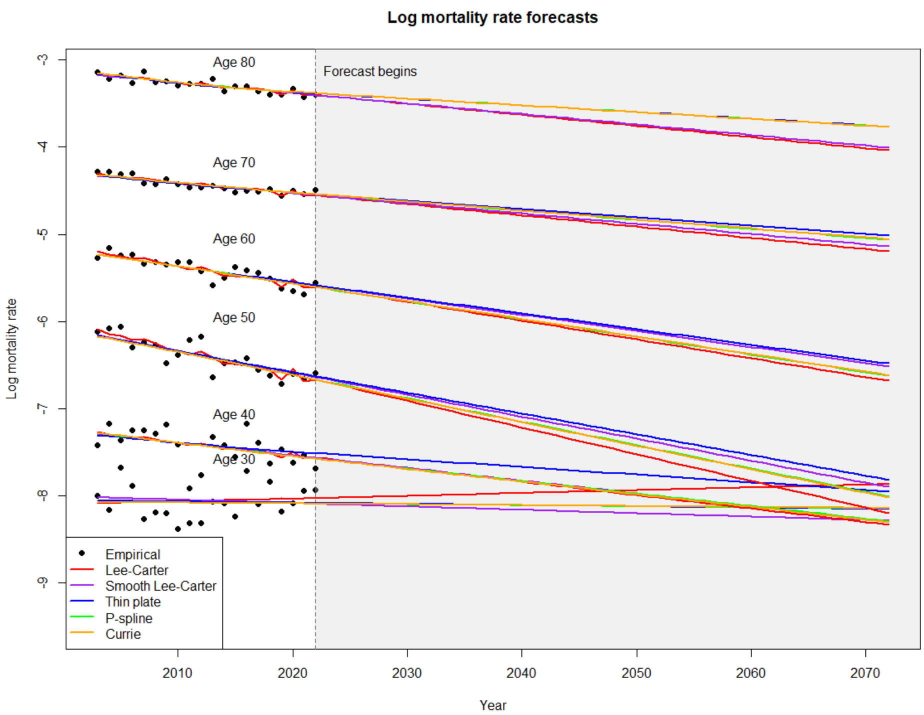


Figure 14: 50-year forecasts for female mortality