A Supervised Mortality Learning

A thesis submitted for the degree of

Bachelor of Commerce (Honours)

by

Zhen Khang Lincoln Sim

30437393

Department of Econometrics and Business Statistics

Monash University

Australia

October 2022

A SUPERVISED MORTALITY LEARNING

Supervised by: Assoc. Prof. Dan Zhu
Department of Econometrics and Business Statistics,
Monash University
Victoria, Australia
Dan.Zhu@monash.edu

Zhen Khang Lincoln Sim
Department of Econometrics and Business Statistics,
Monash University
Victoria, Australia
zsim0003@student.monash.edu

ABSTRACT

Mortality forecasting is one of the fundamental topics in the Actuarial community as it gives analysts an idea about the exposure of mortality risks in the future. The publication of the Lee-Carter (1992) model is a memorable achievement for Actuarial Studies and often serves as a benchmark model for actuaries. In this paper, we introduce the Spline Dynamic Factor (SDF) model which is an extension of the Lee-Carter (1992) model with an application to the French mortality data. A data-driven traditional mortality study automatically selects the factors by an algorithm. However, with information and expert knowledge, we know that a few critical age groups drive the dynamic patterns of mortality experiences. With selections of factors based on it, the proposed model's factors will have a more straightforward interpretation whereas traditional studies are hard to interpret. The forecasts we obtained show an improved performance relative to the Lee-Carter (1992) model and a few of its variants.

Keywords Lee-Carter model · Long-term forecasting · Supervised mortality studies · Robust mortality forecasts · Spline Dynamic Factor Model

1 Introduction

When the word "mortality rate" is mentioned, what comes to mind? Pictures of the mortality table labelling each column with different notations such as age, year, number of deaths, and number of lives? The mortality rate of an age x at time t is denoted by m_{xt} and it is defined as the number of deaths in a particular age group (x) in year t. From the past few centuries, the life expectancy of human beings had increased significantly and it is undeniably a great advancement to the whole human race. However, the advancement in longevity has also brought up another issue which is the longevity risk as well as the mortality risk. Longevity risk is defined as the risk that an individual lives longer than expected where the mortality risk is defined as the risk that an individual dies earlier than expected. Therefore, the projections of human mortality have been an important and frequently discussed topic in various fields of academia and industry, such as Actuarial Science, Time Series, Forecasting, Life Insurance, Annuities Providers, Government and so on. The history of mortality modelling can be traced back to 19^{th} century when the well-known Gompertz Law of Mortality (Gompertz, 1825) and Makeham's law of Mortality (Makeham, 1867) had been published. Generally, life insurance companies and annuities providers are concerned about whether there is a more powerful model that is able to produce a more accurate and reliable mortality projections so that they can reduce the potential loss that arise from the mortality risk and longevity risk respectively when underwriting their products.

In a large course of mortality modelling history, the technique of factor modelling is often used by researchers where there exist some time-variant latent factors that capture the time effects and also some time-invariant factor loadings that capture the relationship between the latent factors and the dependent variables (which generally will be the mortality of an age cohort). The publication of the Lee-Carter model (R. D. Lee & Carter, 1992) hit the milestone of application on mortality modelling and simulated the interests of researchers to explore the mortality phenomenon based on the technique of factor modelling.

Although the LC model has been published for almost 30 years, it is still widely used nowadays as a benchmark model. This is because the model itself is parsimonious, it is easy to apply, and it is easily understood. The LC model contains one time-variant latent factor and one time-invariant factor loading. The model suggests that the mortality experiences across different age groups at time t can be explained by the same and only factor and therefore it is often referred to as the one-factor model. Throughout the 30 years, researchers had found out the limitations of the LC model, extensions, and model reviews had been done to target the downsides of the LC model.

Booth, Maindonald, and Smith (2002) first showed that, by increasing the number of factors, the model performance can be improved. Based on this finding, researchers conducted their research as an extension of the multi-factor mortality model, for example, Yang, Yue, and Huang (2010) applied it to a few countries with gender-specific comparisons. French and O'Hare (2013) applied a macroeconomic forecasting technique and proposed a dynamic factor model of mortality learning that is able to produce superior forecasts. One of the downsides of the LC model is that it does not capture cohort effects. Renshaw and Haberman (2006) then proposed a cohort extension of the LC model and they showed that when considering the relationship between the cohort effect and the mortality experience of each age groups, the model performance can be improved. However, due to the identification issue that arose from the model, Kuang, Nielsen, and Nielsen (2008) proposed a canonical parameterization that is able to solve the identification issue.

Another arguable benchmark model that researchers often refer to is the functional data model proposed by Hyndman and Ullah (2007). The model treats the mortality data as functional time-series data and produced more accurate long-term forecasts that outperform the model proposed by R. Lee and Miller (2001). As technology has improved, different modeling methods have been developed, especially in the Machine Learning field, such as classifications and regression trees (Breiman, Friedman, Olshen, & Stone, 2017), random forests (Breiman, 2001), and neural networks (McCulloch & Pitts, 1943). In this context, Richman and Wüthrich (2021) extended the LC model to multiple populations via the use of a neural network. Besides, Cairns et al. (2009) and Booth, Hyndman, Tickle, and De Jong (2006) listed some comparisons among the variants of the LC model as well as variants of the Cairns-Blake-Dowd (CBD) model. Tuljapurkar, Li, and Boe (2000) then explored the behavior of mortality experience across the years for each of the G7 countries using the LC model.

As mortality data sets are generally high-dimension (i.e. number of variables is much larger than the number of observations) and dimension reduction analysis will be needed to do mortality modelling. Some unsupervised learning methods that are often used to reduce the high-dimensionality includes the Shrinkage Methods via Ridge regression (Hoerl & Kennard, 1970) and least absolute shrinkage and selection operator (LASSO) ((Santosa & Symes, 1986), (Tibshirani, 1996)), the Principal Component Analysis (Pearson, 1901), and so on. The LC model uses the Singular Value Decomposition (SVD) via Principal Component Analysis (PCA). Besides the frequentist paradigm, researchers also extend their studies on morality forecasts to the Bayesian paradigm and recent papers include but are not limited to Lazar and Denuit (2009), Li, Li, Tan, and Tickle (2019), and Zhu, Shi, Shi, and WANG (n.d.).

Gaps in the literature on the LC model include its limited range of years for the US data and also the potential overfitting problem that arose from the multi-factor mortality modelling. Currently, available pieces of literature are building its block on the LC model and focusing on the data-driven model, which is considered as an unsupervised learning mortality learning. It assumes that there is no relationship between the response variable (i.e. the mortality rate) and the independent variable (i.e. the age groups) where factors given in the model are decided by the data itself. Moreover, factors selected by the data when conducting unsupervised mortality learning, do not have much interpretability and they lack economic meaning. Nonetheless, the fact is that the mortality rate does have a dynamic pattern and the dynamic pattern is driven by a few important factors (i.e. age groups).

Besides, since Booth et al. (2002) proposed their findings on multi-factor mortality modelling can improve forecast performance, researchers and actuaries had been applying these findings to their own mortality studies. By increasing the number of factors, the model is able to capture more information on the mortality data and so it is able to produce a better model fit as well as forecast performance. However, there exists a turning point where the forecast performance will not improve anymore when the number of factors increases although the model fit will keep improving. This is because the model overfits the training data.

Therefore, our research question will then be, "By controlling the complexity of our model, and conducting a supervised mortality learning with factors selected priory, can we improve the model performance?". In light of these, we propose the Spline Dynamic Factor model (SDF) to fill the gaps. We introduce a supervised mortality learning where we are able to select some factors that have economic meaning ourselves. By introducing the SDF model, we aim to improve the forecast performance by reducing the chance of overfitting.

The rest of the paper is organised as follows. In Section 2, we introduce the model specification of our proposed model. The Section makes starts by briefly refreshing the reader's mind with the milestone model of mortality modelling, the LC model, which is a classical one-factor model. Not far from the LC model being introduced, researchers found out that by increasing the number of factors, the model performance can be improved. Therefore, Section 2 then introduces the Extended LC model which is a classical multi-factor model that extends the LC model by adding more factor that helps the model to capture the mortality experiences. The section then introduces the identification scheme as well as the estimation procedure of the proposed model.

Section 3 then discusses the empirical findings and analysis starting by first exploring the French male mortality data and a model performance investigation has also been conducted in this section through the in-sample fit and the out-of-sample fit and a forecasting method has been finalized as well. The section then lists out each of the ARIMA models fitted to each latent common factor and illustrates the empirical findings of the latent common factor. It shows that the latent common factor captures the mortality experience well and a supervised mortality learning method should be considered in the future. Section 4 then concludes the paper.

2 Model Specification

Let m_{xt} denote the mortality rate for age x in year t, where $x = 0, 1, 2, ..., X$ and $t = 1, 2, ..., T$. Thus the observed mortality data **M** is an $(X + 1) \times T$ matrix. A log-transformation of the mortality data is used to remove the exponential rise.

2.1 The Classical Factor Model

The actuarial community has been long interested in the study of mortality rates, starting from the seminal work of (R. D. Lee & Carter, 1992). They specify

$$log(m_{x,t}) = \boldsymbol{\alpha_x} + \boldsymbol{\beta_x \kappa_t}$$

where $\boldsymbol{\kappa_t}$ is the univariate driving force for the time series. While the original LC model was introduced as a deterministic model, later extensions of the model to include an error term

$$log(m_{x,t}) = \boldsymbol{\alpha_x} + \boldsymbol{\beta_x \kappa_t} + \boldsymbol{\epsilon_{x,t}}$$

as well as a state-space formulation of $\boldsymbol{\kappa_t}$

$$\boldsymbol{\kappa_t} = b + B\boldsymbol{\kappa_{t-1}} + z_t$$

for forecasting purposes.

Even since its introduction, it has been the workhorse model in mortality modelling and understanding of basis risks in life insurance contracts. There are many criticisms of the model, surrounding the overly simplified framework of having a univariate temporal dynamic. This limitation of the LC model motivated a plethora of extensions, i.e., the application of the classical factor model to mortality studies.

The classical factor model application defines the extension of the LC model as:

$$log(m_{x,t}) = \boldsymbol{\alpha_x} + \sum_{k=1}^{K} \boldsymbol{\beta_{x,k}} \boldsymbol{\kappa_{k,t}} + \boldsymbol{\epsilon_{x,t}} \tag{1}$$

where $\boldsymbol{\alpha_x}$ and $\boldsymbol{\beta_{x,k}}$ are the age-specific parameters, $\boldsymbol{\kappa_{k,t}}$ is the time effect parameter, and the error term $\boldsymbol{\epsilon_{x,t}}$ is assumed to be identically independent distributed (i.i.d) across all age cohorts with a mean of 0 and a variance of $\sigma^2_{x,t}$.

The classical factor model first estimates $\boldsymbol{\alpha_x}$ by simply taking the average of $log(m_{x,t})$ across all time periods, which is

$$\boldsymbol{\hat{\alpha}_x} = \frac{1}{T} \sum_{t=1}^{T} log(m_{x,t}).$$

Then a mean adjusted mortality data will be modelled by the Singular Value Decomposition (SVD). The estimated $\boldsymbol{\hat{\alpha}_x}$ is a $(X + 1) \times 1$ vector, $\boldsymbol{\hat{\beta}_{x,k}}$ is a $(X + 1) \times K$ matrix and $\boldsymbol{\hat{\kappa}_{k,t}}$ is a $K \times T$ matrix.

The notation k in equation 1 refers to the number of factors where it has a range of values from $1, ..., K$. When $K = 1$, the model will then indicates the one-factor model that was proposed by R. D. Lee and Carter (1992) (i.e. The LC Model) and we will be calling it the "Standard LC" model in the rest of our paper. When $K > 1$, the model will then indicates the multi-factor model that was first shown by Booth et al. (2002) of its ability to explain more of the mortality information and we will be calling it the "Extended LC" model in the rest of the paper.

Due to the nature of the defined model, the forecast of the system is thus driven by the latent factor $\boldsymbol{\kappa_{k,t}}$. The LC model (1992) (when $K = 1$) allows the state-space equation to follow a Random Walk with Drift (RWD) process. The h step ahead forecasts given historical mortality data up to time t is obtained by

$$log(m_{x,t}) = \boldsymbol{\alpha_x} + \boldsymbol{\beta_x} \boldsymbol{\hat{\kappa}_{t+h}} + \boldsymbol{\epsilon_{x,t}}$$

where $\boldsymbol{\hat{\kappa}_{t+h}} = h\hat{\mu} + \boldsymbol{\kappa_t}$ and $\hat{\mu} = \frac{1}{n-1}\sum_{t=1}^{n-1}(\kappa_{t+1} - \kappa_t)$. Besides of the traditional RWD process, researchers such as Hyndman and Ullah (2007), had also allowed the state-space equation to follow an Auto-Regressive Integrated Moving Average (ARIMA) process.

Another forecast method used by researchers such as Lazar and Denuit (2009) is the Vector Auto-Regressive (VAR) process and it is a method commonly used by the macroeconometrics field where we further the discussion on the application of the VAR process on mortality forecasting in the later section.

2.2 The Spline Dynamic Factor Model

As mentioned in Section 1, current mortality studies are built on unsupervised learning and the fact is mortality data does have a dynamic trend. We introduce the Spline Dynamic Factor (SDF) model that extends the classical factor model in various ways.

The SDF model is built on the Smoothing Splines that were introduced by Green and Silverman (1993) and the dynamic factor mortality model that was introduced by French and O'Hare (2013). The short of the Smoothing Splines technique is that we do not have to worry about any changes in the number of knots chosen and placed will lead to huge a impact on the model performance as it uses all of the data points as an individual knot. To overcome the potential over-fitting due to a huge number of knots, the Smoothing Splines method introduces a penalty term where we will be discussing the estimation of the SDF model in the relevant section.

We consider mortality rates from age $x \in \mathcal{X} = [0, ..., X]$ during the year $t \in \mathcal{T} = [1, ..., T]$. We define our measurement equation as:

$$log(m_{x,t}) = \boldsymbol{\alpha(x)} + \sum_{k=1}^{K} \boldsymbol{\beta_k(x)\kappa_{k,t}} + \boldsymbol{\epsilon_{x,t}}, \boldsymbol{\epsilon_{x,t}} \sim N(0, \sigma_x^2) \tag{2}$$

where the $(X+1) \times 1$ vector $\boldsymbol{\alpha(x)}$ and the $(X+1) \times K$ matrix $\boldsymbol{\beta(x)}$ are a smooth functions of age. The factors $\boldsymbol{\kappa_t} \in \mathcal{K} \subset \mathbb{R}^K$ follow

$$\boldsymbol{\kappa_t} | \boldsymbol{\kappa_{t-1}}, ...\boldsymbol{\kappa_{t-p}} \sim F.$$

and the exact specification of the latent dynamics factor depends on the application. We shall discuss this further in the empirical section.

The SDF model differentiates from existing models from a few perspectives. Firstly, it is a supervised learning method where K-selected factors based on the prior knowledge that the overall dynamics of the mortality experience are driven by them can be targeted. This is a reasonable proposed model that benefits from the nature of the identification scheme which we will be discussing in a later section. Secondly, the model uses the few major factors that drive the dynamics pattern of the mortality experience are able to "learn" the whole mortality pattern without "learning" much about the underlying noises.

Thirdly, due to the prior selection of which factor to use instead of letting the data decide, the SDF model provides plenty of interpretability results as well as robustness which we will be discussing them in the empirical analysis section. Lastly, it provides modellers flexibility when applying the SDF model to other populations as they are able to select different factors based on the demography and economic status of the area they are examining the mortality behavior.

2.2.1 Identification Scheme

As mentioned in Section 1, the Standard LC model (i.e. one factor classical model) is subject to common identification issues. The identification scheme used by R. D. Lee and Carter (1992) is to restrict the loading factor is added up to one ($\sum_x \boldsymbol{\beta_x}$=1) and the latent common factor is added up to zero ($\sum_t \boldsymbol{\kappa_t}$=0).

For multiple factor models, Bai and Wang (2015) rewrite the model in its static factor form,

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta\kappa_t} + \boldsymbol{\epsilon}_t, \;\; \boldsymbol{\epsilon}_t \overset{i.i.d}{\sim} N(\mathbf{0}, \sigma^2\mathbf{I})$$

where $\mathbf{y}_t = [log(m_{0,t}), ..., log(m_{X,t})]'$, $\boldsymbol{\alpha} = [\alpha_0, ..., \alpha_X]'$, and the $(X+1) \times K$ matrix $\boldsymbol{\beta}$ is the factor loading for the $K \times T$ matrix $\boldsymbol{\kappa_t}$ which is the latent factor.

We further discuss specification of the latent dynamics factor mentioned in Section 2.2. The latent factor follows a VAR(h) process

$$\boldsymbol{\kappa}_t = \mathbf{b} + \sum_{j=1}^{p} \mathbf{B}_j \boldsymbol{\kappa}_{t-j} + \mathbf{z}_t, \quad z_t \overset{i.i.d}{\sim} N(0, \Omega_K)$$

where $\mathbf{b}$ is a $h \times 1$ vector of the intercept and $\mathbf{B}$ is a $K \times K$ matrix of autoregressive coefficients. By assuming the latent process at time 0 is 0 (i.e. $\boldsymbol{\kappa}_0 = 0$), the factor will then follow

$$\boldsymbol{\kappa_t} | \boldsymbol{\kappa_{t-1}}, ...\boldsymbol{\kappa_{t-p}} \sim N(b, \Omega_K).$$

Due to the rotational indeterminacy of factor models, they are not identifiable without further restrictions and so two identification scheme are suggested by them. Let

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{bmatrix},$$

where $\boldsymbol{\beta}_0$ is $K \times K$ matrix and $\boldsymbol{\beta}_1$ is then a $(X+1-K) \times K$ matrix.

The two suggested identification schemes are:

1. var(Ω_K)=$\mathbb{I}_K$ and the $K \times K$ matrix $\boldsymbol{\beta}_0$ is restricted to be lower-triangular matrix with strictly positive diagonal elements.

2. Imposing restriction only on the factor loading, which is restricting $\boldsymbol{\beta}_0$ to be an identity matrix.

In our paper, we will be using the second set of identification schemes for the SDF model. In this particular set of identification schemes, it is only restricting the factor loading and so the VAR(h) process latent factor is completely unrestricted. For more details and proofs, please refer to Bai and Wang (2015).

Once we fix the identification of the model by setting $\boldsymbol{\beta}_0 = \mathbb{I}_K$, it implies that the first K variables in the system evolves according to

$$\mathbf{y}_{t,1:K} = \kappa_t + \epsilon_{t,1:K}$$

that is the first K variables drives the overall dynamics of the system. Hence, placing which variables in in the first K elements of the vector becomes critical. Yet, for majority of applications in practice, modellers would already have some prior knowledge on the structure of mortality rate.

The SDF model has the ability to target those factors that modellers select by the structure of the model. For example, we consider targeting five age groups which are age 15, age 20, age 25, age 30, and age 35. When constructing a model with four factors, the classical factor model orders the mortality data in chronological order of age. However, the SDF model first shuffles the five targeted age groups to the first five rows of the data and so the mortality data will be in such an order of $x = 15, 20, 25, 30, 35, 0, 1, 2, ..., X$. The four factors SDF model will then be targeting the top four age groups which are age 15, age 20, age 25, and age 30.

2.3 Estimate the SDF Model

We shall discuss the exact iterative estimation procedure that will be done iteratively to obtain relevant parameters for the SDF model. Let $\mathcal{A}_0$ denote the selected age placed in the first K elements of $\mathbf{y}_t$ for identification purposes, and $\mathcal{A}_0^c = \mathcal{X} - \mathcal{A}_0$.

- Given the factors and the parameter of the latent model, we estimate the age specific coefficients. The likelihood of the coefficients can be written as:

$$\underset{\alpha\in\mathcal{A},\beta\in\mathcal{B}}{argmax} \sum_{x\in\mathcal{A}_0^c}\sum_{x=1}^{X} -\frac{T}{2}log(\sigma_x^2) - \frac{1}{2\sigma_x^2}\sum_{t=1}^{T}[log(m)_{x,t} - \boldsymbol{\alpha}(x) - \sum_{k=1}^{K}\boldsymbol{\beta}_k(x)\boldsymbol{\kappa}_{x,t}]^2$$

  where $\mathcal{A}$ and $\mathcal{B}$ are the set of $\boldsymbol{\alpha}(x)$ and $\boldsymbol{\beta_k}(x)$. As mentioned in Section 2.2.1, both $\boldsymbol{\alpha}(x)$ and $\boldsymbol{\beta}(x)$ are a smooth function of age where the function can be in polynomial form, quadratic form, B-spline, and the list go on. We consider the smooth function to be the Smoothing Splines that was introduced by Green and Silverman (1993). To incorporate the Smoothing Splines, we considering optimising

$$\underset{\alpha\in\mathcal{A},\beta\in\mathcal{B}}{argmax} \sum_{x\in\mathcal{A}_0^c}\sum_{x=1}^{X} -\frac{T}{2}log(\sigma_x^2) - \frac{1}{2\sigma_x^2}\sum_{t=1}^{T}[log(m_{x,t}) - \alpha(x) - \sum_{k=1}^{K}\beta_k(x)\kappa_{k,t}]^2 + \lambda_1\int\alpha''(t)dt + \lambda_2\int\beta_k''(t)dt$$

- The choice of λ_1 and λ_2 will be chosen respectively by minimizing:

$$RSS_{cv}(\lambda) = \sum_{i=1}^{X}[\frac{y_i - \hat{g}_\lambda(x_i)}{1 - \{\boldsymbol{S}_\lambda\}_{ii}}]^2$$

  where $\lambda = \{\lambda_1, \lambda_2\}$, $y_i = \{\alpha, \beta\}$, $\hat{g}_\lambda(x_i)$ denotes the smooth function of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ at x_i, and $\{\boldsymbol{S}_\lambda\}_{ii}$ denotes the effective degrees of freedom. There are different ways that the above minimization can be done and in our paper, we will be using the leave-one-out cross-validation. For more details, please refer to Green and Silverman (1993).

- Given the age specific coefficients and the parameter of the latent model, estimate the factors

$$\underset{f}{argmax} \sum_{x\in\mathcal{A}_0^c}\sum_{x=1}^{X}\frac{1}{2\sigma_x^2}\sum_{t=1}^{T}[log(m_{x,t}) - \alpha(x) - \sum_{k=1}^{K}\beta_k(x)\kappa_{k,t}]^2 + \sum_{t=p+1}^{T} log[f(\boldsymbol{\kappa}_t|\boldsymbol{\kappa}_{t-1}, ..., \boldsymbol{\kappa}_{t-p}, \boldsymbol{\theta})]$$

  where $\boldsymbol{\theta}$ denotes the parameter of the latent model, $\boldsymbol{\theta} = [\mathbf{b}, \mathbf{B}, \Omega_K]$.

- Given the age specific coefficients and the factors, estimate the parameter of the model $\boldsymbol{\theta}, \sigma_x^2$

$$\underset{\sigma_x^2,\boldsymbol{\theta}}{argmax} \sum_{x\in\mathcal{A}_0^c}\sum_{x=1}^{X}\frac{1}{2\sigma_x^2}\sum_{t=1}^{T}[log(m_{x,t}){-}\alpha(x){-}\sum_{k=1}^{K}\beta_k(x)\kappa_{k,t}]^2{+}\sum_{x=0}^{X}log[f(\boldsymbol{\kappa}_{t-1},...,\boldsymbol{\kappa}_{t-p}|\sigma_x^2,\boldsymbol{\theta})]$$

3 Empirical Findings and Analysis

3.1 Data

The observed French male mortality data will be extracted from the Human Mortality Database (HMD) [1], from age 0 to 110+ from 1816 through. Due to the unreliability of observing mortality data, we will then be using age 0 to 100 (i.e. 101 age groups (X=101) in total and x = 0,1,...,100) .

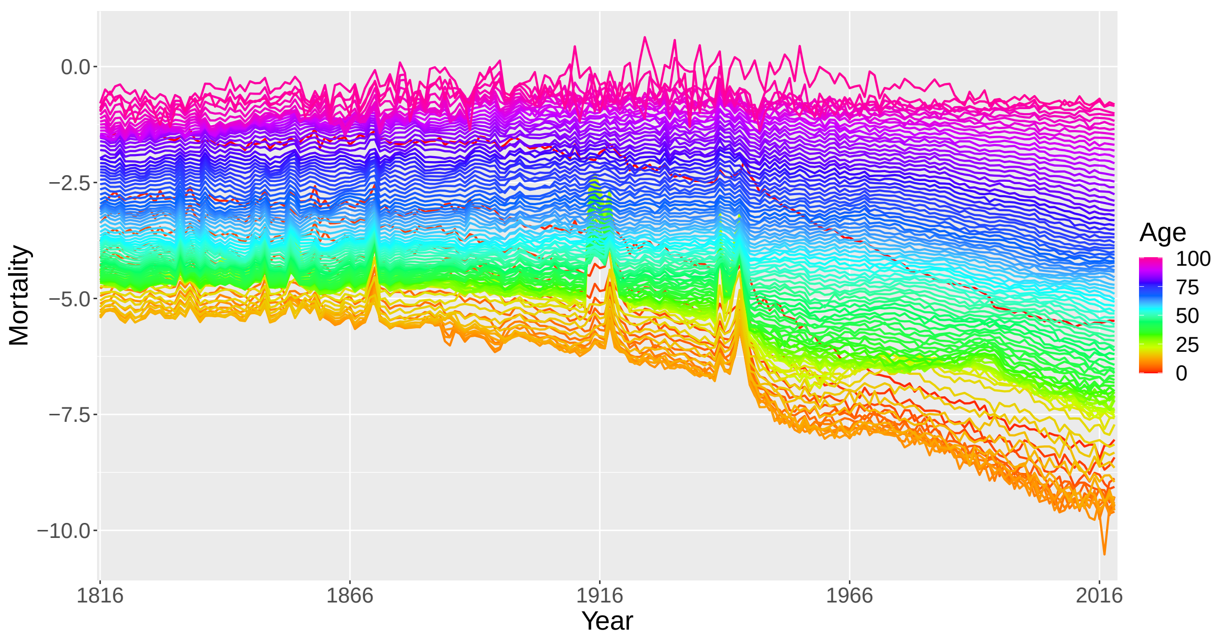

Figure 1: The above plot illustrates $log(m_{xt})$ for each age groups (age 0 to age 100) across the years (1816-2019).

Figure 1 illustrates the mortality experience of French male age 0 to age 100 from 1816 to 2019. It shows an overall downward trend although the decline rate differs across the age groups. By observing from Figure 1, we can see that the overall dynamic pattern of mortality experience for the past few decades is driven by a few variables. Hence, based on prior knowledge of the structure of the mortality rate, the SDF model is able to capture the whole mortality pattern instead of capturing the noise.

As we all know, World War I happened from 1914 to 1918 after being triggered by the assassination of Austrian Archduke Franz Ferdinand, and World War II happened from 1939 to 1945 after the Invasion of Poland by Adolf Hilter. Mortality data are very sensitive to extreme events from Figure 1, the impacts of World War I and World War II are very significant and can be easily

[1]HMD : `https://former.mortality.org/hmd/`

seen in the figure. Since the use of force in international conflicts is prohibited by international law, we decided to use data from 1950 onwards as such a worldwide war rarely happens.

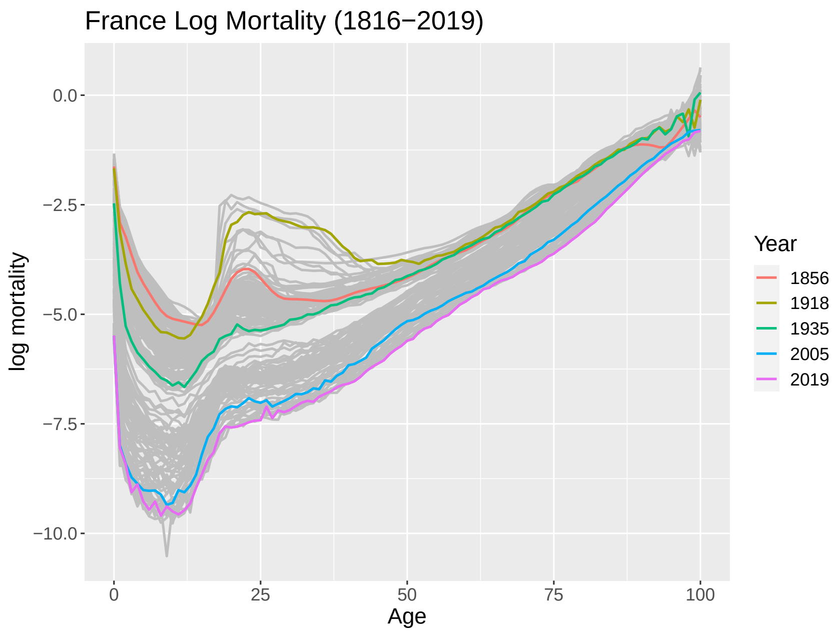

Figure 2: This shows a significant decrease of log mortality rates across the years.

As technology improves, people enjoy a more comfortable lifestyle as well as a better medical system and they maintain a higher level of hygiene that helps to improve mortality more than expected (i.e. We see a significant decrease in mortality rates, as shown in Figure 2). From Figure 2 we are able to observe that throughout the year, there is always a significant decline after the infant mortality rate. The unusual humps during the teenage year are often referred to as the "accident hump" as the main reason of it is accidents. However, there seems to exist an outlier year that is indicated by the yellow line and that was due to World War I and the Spanish Flu.

3.2 Factor Selection

We begin with factor selection, where we select some important age groups that drive the dynamic pattern of the mortality rates based on some prior knowledge of the structure of the mortality rate. The first age group that we select is age 0, which the mortality rate often refers to as the infant mortality rate. Causes of infant death in France include socio-economic status, closure of many maternity hospitals, and Sudden Infant Death Syndrome (SIDS) which France leads the SIDS rate among European Union (EU) countries. Researchers and Actuaries often refer to the mortality rates of infants as well as older age mortality rates as a proxy of the development of a country, therefore we will be selecting age 99 as our second factor.

Thirdly, we selected age 33 as the third important age. The global pandemic of HIV/AIDS (Human Immunodeficiency Virus Infection and Acquired Immune Deficiency Syndrome) began in 1981 after the first ever cases were reported in the United State of America (U.S.A) on the 5^{th} of June 1981. Since the declaration of the HIV/AIDS global pandemic, people from age 15 to age 49 are leading the death count. Age 33 is selected to capture the impact of the HIV/AIDS global pandemic on French male mortality rates as it is the mean age of the age group 15 to 49. We then consider age 21 as the fourth important age. In France, age 21 is generally the age when one graduates from university and just gets into the workforce where a totally different lifestyle is adopted. Age 21 is also the age that one has the capability to perform all the act's of civil life.

We then move on to the fifth selection of important factors and age 66 is selected. The average retirement age in the EU Member States is 65 while France's retirement age is 62. However, the majority of the EU Member States have decided to raise the retirement age to 67. We believe that France will raise their retirement age to at least above 65 and so age 66 is selected. Lastly, we consider age 18 as the next important age. According to the World Population Review 2022, the legal age to purchase and consume all alcoholic beverages is 18 and it is also the legal age in France to drive. Subtle differences between having age 18 and age 21 as factors are that we believe that when an individual has the legal right of consuming alcohol and drive, it will have an impact on the mortality rate at age 18. When an individual has to adopt a different lifestyle, the associated mental issues might impact the mortality rate at age 21. Due to the significant accident hump for the past few centuries, we think that it might be helpful if we use two factors to capture that mortality experience.

Based on prior experiences from multi-factor mortality model-related literature, the optimum number of factors will be around five and six. Therefore, we decided to select the six most important factors and the best SDF model might have six factors or less. In summary, based on prior knowledge and expert opinion, we have selected age 0, age 99, age 33, age 21, age 66, and age 18 as six of the most significant factors that drive the whole dynamic pattern of mortality experiences.

3.3 Model Performance

We use the Mean Squared Error[2] to measure the goodness of fit and the forecast accuracy of the model. A time series cross-validation will be applied where a base period will be used for model fitting, and the rest of the series of data will be used to examine the forecast accuracy of each model.

We first decide that the base period to be from 1950 to 2000 and we will then roll the window as we expanding it. For example, when the forecast horizon is 5 (i.e. h=5), the training data set from 1950 to 2000 will be used to train the model and it will be used to forecast 2005. The in-sample MSE will be computed based on the training data and the out-of-sample MSE will be computed based on the testing data.

After both the in-sample and out-of-sample MSE are obtained, we then expand our training data set one step ahead to 2001, model will be trained using data from 1950 to 2001, and forecast accuracy will be determined by 2006. The last training data set will be from 1950 to 2014 and the last testing data set will be 2019. The overall goodness of fit and the overall forecast accuracy for $h = 5$ will be computed by averaging over the 14 collected in-sample MSE and out-of-sample MSE. This is normally referred to as the "Rolling-Window Backtesting" and it is often used by

[2]The MSE is defined as $MSE = \frac{1}{T}\frac{1}{X}\sum_{t=0}^{T}\sum_{x=0}^{X}[\epsilon_{x,t}]^2$ where $\epsilon_{x,t} = log(\hat{m}_{xt}) - log(m_{xt})$

researchers to examine the model performance and forecast accuracy, such as in Tang, Li, and Tickle (2022).

We will be comparing the model performance of the SDF model along with the Standard LC model and the Extended LC model. We will also consider the functional data model that was proposed by Hyndman and Ullah (2007) and we will be calling it the "Hyndman_Ullah" model in the rest of the paper. The Hyndman_Ullah model is an extension of the Standard LC model (i.e. the classical factor model when $k = 1$) and it is also served as another benchmark model due to the superior long-term forecast results that outperform R. Lee and Miller (2001) methods. The order K of the Hyndman_Ullah model is determined by the Integrated Squared Forecast Error. For more details, please refer to Hyndman and Ullah (2007).

3.3.1 In-Sample Fit

Forecast Horizon	Number of Factor	Hyndman_Ullah	Extended LC	SDF
1	1	0.0180	**0.0087**	0.0162
	3	0.0047	**0.0035**	0.0123
	6	0.0029	**0.0023**	0.00672
5	1	0.0146	**0.0084**	0.0143
	3	0.0040	**0.0034**	0.0119
	6	0.0026	**0.0022**	0.0066
10	1	0.0118	**0.0080**	0.0124
	3	0.0037	**0.0033**	0.01122
	6	0.0025	**0.0022**	0.0063
15	1	0.0100	**0.0076**	0.0107
	3	0.0035	**0.0033**	0.0101
	6	0.0023	**0.0021**	0.0066

Table 1: In-Sample factor-specific h-step MSE by method. (Hyndman-Ullah: Functional Data Model in Hyndman and Ullah (2007), Standard LC: The standard LC model in R. D. Lee and Carter (1992), Extended LC: The classical factor model with $K > 1$, SDF: The Spline-Dynamic Factor model described in Section 2.2))

We present the results of in-sample fit for the four models across the forecast horizons in Table 1. We see that the Extended LC model dominates the in-sample performance with a very large margin, as the number of factors increase, the MSE decreased for all forecast horizons. The model performance of the Extended LC model is at least twice larger than the SDF model. The Hyndman_Ullah model also performs relatively better than the SDF model and the difference between it and the Extended LC model does not vary much. Therefore, both the Hyndman_Ullah model and the Extended LC model work better when it comes to fitting the training data. The SDF model performs the worst across the time horizons among the three models which are expected.

Improvement of in-sample fit can always be obtained by simply increasing the number of factors. However, the potential issue that arise when using such a complex model to capture the mortality experience are that it will overfit the training data which leads to a bad out-sample performance which we will show in the next subsection.

3.3.2 Out-Of-Sample Fit

We showed in Section 2.2.1 the form of the latent factor when it follows a VAR(h) process and taking into consideration of overparameterization problem, the latent factor will then be following a VAR(1) process. Besides of the VAR(1) process, the ARIMA model is often used when doing time-series forecast and application of the ARIMA model on mortality forecasting is very common as well. Therefore, we consider the ARIMA model by allowing the latent factor to follow an ARIMA(p,d,q) process where it can be shown as

$$(1-\phi_w B)(1-B)\boldsymbol{\kappa_t} = c + (1+\gamma_z B)e_t \;\; e_t \overset{i.i.d}{\sim} N(0, U_K)$$

where ϕ_w is the coefficient of the AR part, γ_z is the coefficient of the MA part, c is a constant, and B is a Back-Shift notation[3]. When considering an ARIMA process, the latent model parameter mentioned in Section 2.3 will then be $\boldsymbol{\theta} = [\phi_w, \gamma_z, U_K]$ and the order of the ARIMA process (i.e. p, d, q) will be determined by a suitable model selection criteria which is AICc and the reason will be discussed in Section 3.4.1.

Denote $\mathcal{F}_t$ as the observed data $\mathcal{F}_t = \{log(m_{xt}); t = 1, ..., T; x = 0, ..., X\}$. With equation 2 and one of the suitable state equation, the h steps ahead forecasts conditioned on $\mathcal{F}_t$ can be obtained by:

$$log(\hat{m}_{x,t+h}) = E[log(m_{x,t+h})|\mathcal{F}_t] = \boldsymbol{\alpha(x)} + \sum_{k=1}^{K} \boldsymbol{\beta_k(x)\hat{\kappa}_{k,t+h}}$$

where we denote $\hat{\kappa}_{k,t+h}$ as the h steps ahead forecast conditioned on data observed until time t.

Before connecting our model performance to other approaches, we explore the forecast performance between the two potential methods. From Table 2, we can observe that the ARIMA method is able to produce a more stable forecast than the VAR method, even though the MSE looks very close. We, therefore decided that the state equation of the SDF model will be allowed to have an ARIMA process.

Forecast Horizon	Number of Factor	ARIMA	VAR
1	1	**0.0699**	0.0709
	3	**0.0094**	0.0101
	6	**0.0095**	0.0099
5	1	**0.0859**	0.0962
	3	**0.0169**	0.0271
	6	**0.0175**	0.0213
10	1	**0.1141**	0.1382
	3	**0.0365**	0.0611
	6	**0.0364**	0.0466
15	1	**0.1479**	0.1827
	3	**0.0692**	0.0950
	6	**0.0678**	0.0944

Table 2: A comparison of the SDF model out-of-sample performance between the ARIMA method and the VAR method.

[3] Back-Shift notation is defined as: $By_t = y_{t-1}$

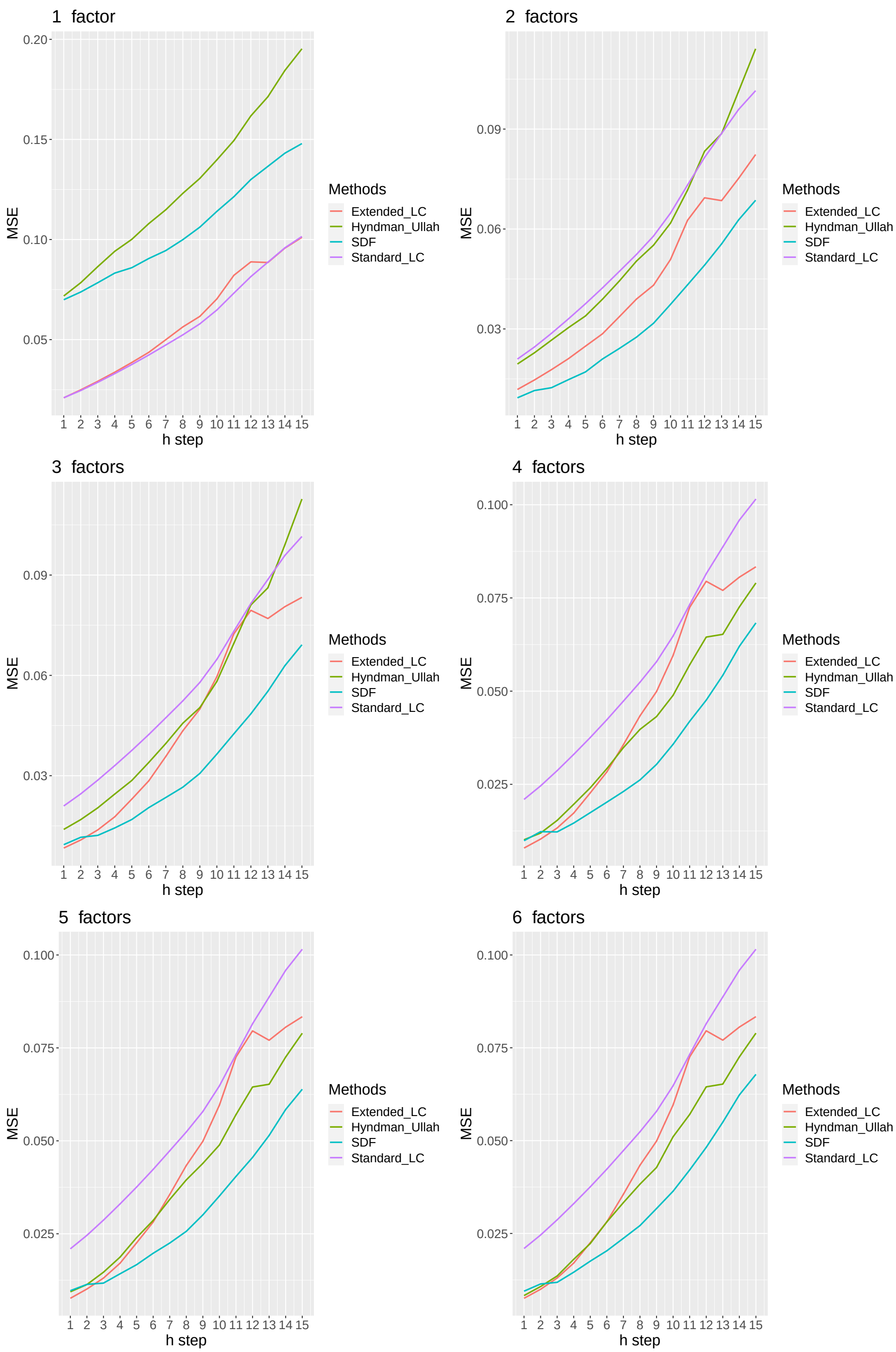

Figure 3: An out-of-sample comparisons of model performance among the four models.

We then present the results of the out-of-sample fit for the four models across the forecast horizons in Figure 3. We include various forecast horizons from 1 to 15 years which allow a comparison of performance for the four models from a short-term perspective to a long-term perspective.

We allow the standard LC model's state equation to follow an RWD process and so it is identical to what was proposed by R. D. Lee and Carter (1992). For the Extended LC model, the Hyndman_Ullah model, and the SDF model, we allow the state equation to follow an ARIMA process. By doing this, we reduce any possibilities that the forecast accuracy is affected by the forecast method that will blur out our aim in this section, which is investigating the differences in forecast accuracy driven by different models.

From Figure 3, we see that when the number of factors is one $(K = 1)$, the Standard LC model performs the best among the four models. When we increase the number of factors, we allow the model to be more complicated so that more mortality information can be captured, there is a huge improvement for the Extended LC model, the Hyndman_Ullah model, and the SDF model.

Forecast Horizon	Number of Factor	Hyndman_Ullah	Extended LC	SDF
1	4	0.0101	**0.0079**	0.0099
	5	0.0094	**0.0076**	0.0097
	6	0.0083	**0.0075**	0.0095
5	4	0.0241	0.0227	**0.0174**
	5	0.0240	0.0226	**0.0167**
	6	0.0223	0.0225	**0.0175**
10	4	0.048885	0.05957	**0.0357**
	5	0.048881	0.05961	**0.0352**
	6	0.0511	0.05961	**0.0364**
15	4	0.0790	0.08336	**0.0683**
	5	0.07894	0.08339	**0.0639**
	6	0.07895	0.08339	**0.0678**

Table 3: Further investigation shows that the optimal number of factors for the SDF model will be five.

We see that when $K > 3$, the Standard LC model performs the worst. The SDF model remains the best when the forecast horizon is greater than two. Since it is difficult for Figure 3 to exhibit the optimal number of factors, we then refer to Table 3 for the selection of the order K. Table 3 shows that the SDF model performs the best when $K = 5$ across the forecast horizons where the omitted factor will be the 6^{th} factor (age 18). The rest of the empirical analysis will be done with the 5-factor SDF model which contains age 0, age 99, age 33, age 21, and age 66 .

3.4 30-Year Forecast

We then produce a 30 years mortality forecast using the 5-factor SDF model. It is shown in Figure 4 where historical data is plotted in grey and 30 years forecasts are colored with rainbow. We can see that the SDF model predicts that the accident hump in teen age will be flatter with significant improvement in mortality rates. On the other hand, from around age 38 to age 100, the SDF model predicts smooth and decline mortality experience for the next 30 years.

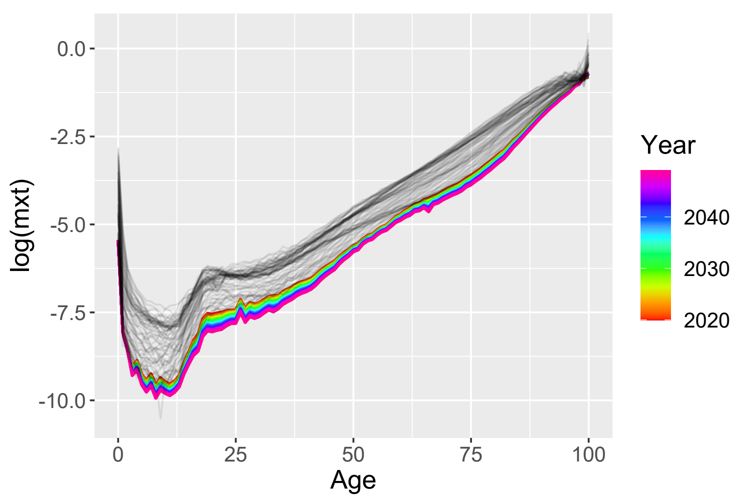

Figure 4: A 30 years mortality forecasts along with log mortality rates from 1950 to 2019.

Given the above mortality forecasting, pensions and life insurance providers have to be aware of the underlying mortality risks for the next 30 years, especially for ages between 0 to 37.

3.4.1 The 5 Fitted Factors

We then examine the five fitted latent factor. The "SDF_fitted" in Figure 5 represents the latent factor from the SDF model where the "Arima_fitted" represents the latent factor that follows an ARIMA process along with the 95% prediction intervals. The dashed lines show the point forecast and the dotted lines show the 95% interval that the chance of future observed value will fall in between the intervals. There is a strong correlation between the SDF fitted latent factors and the mortality experiences that each of the selected important factors targeted.

Since the Akaike Information Criterion (AIC) is well-known for its fairness in providing the goodness of fit of the model but penalizes for overfitting at the same time. Due to the relatively small sample sizes, we consider AIC with another penalty term, which is also known as AICc, to be used in selecting an appropriate ARIMA model for each latent common factor.

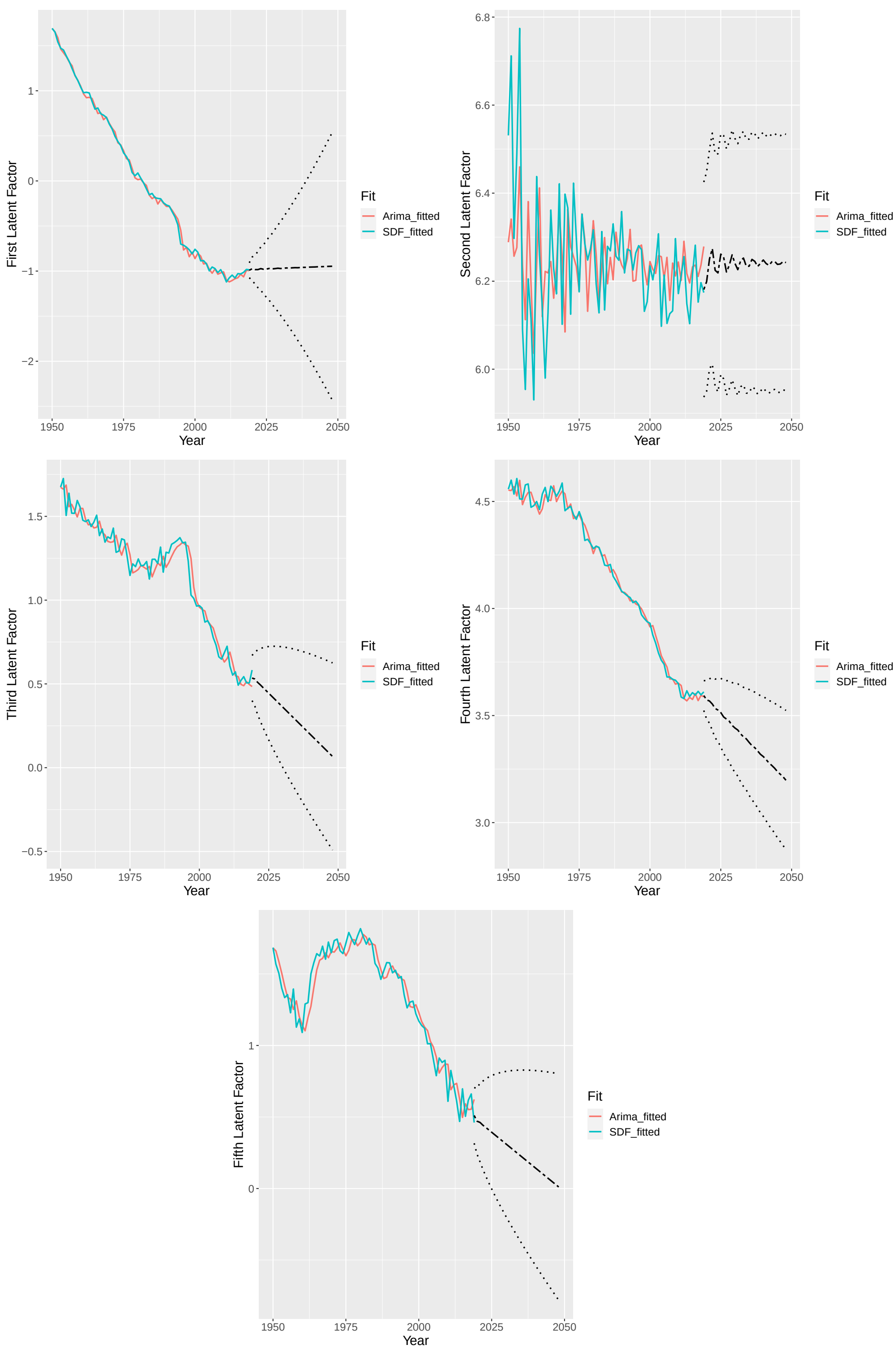

Figure 5: The five latent common factors of the SDF model and the state equation that allowed to follow an ARIMA process.

Figure 5 shows that the ARIMA model is able to follow the pattern of each latent factors. The first and second factor have the widest prediction intervals due to the steepest slope and high variability. It is predicted that in the next 30 years, infant mortality will remain at the same level as well as the old age mortality is predicted to have less variability. However, the associated prediction intervals are the widest among the five fitted factors. The significant “hump” in the third factor indicates the global pandemic of HIV/AIDS and the ARIMA model pick the information up soundly.

We see that each of the latent common factors captures the mortality experience for the corresponding age groups well and these age groups are the main tension to drive the mortality experience of France’s male population. By setting up the SDF model, we restrict each of the factors to pick up a piece of particular information given by an age group’s mortality experience throughout the years. The results tell us that, it is a useful idea and we can therefore say that, a factor-driven supervised learning method, can lead to more precise and accurate results. We have also shown that, for mortality forecasting, a supervised learning method can be applied and an extension can be built on it.

3.4.2 Connection Between Classical Multi-factor Model

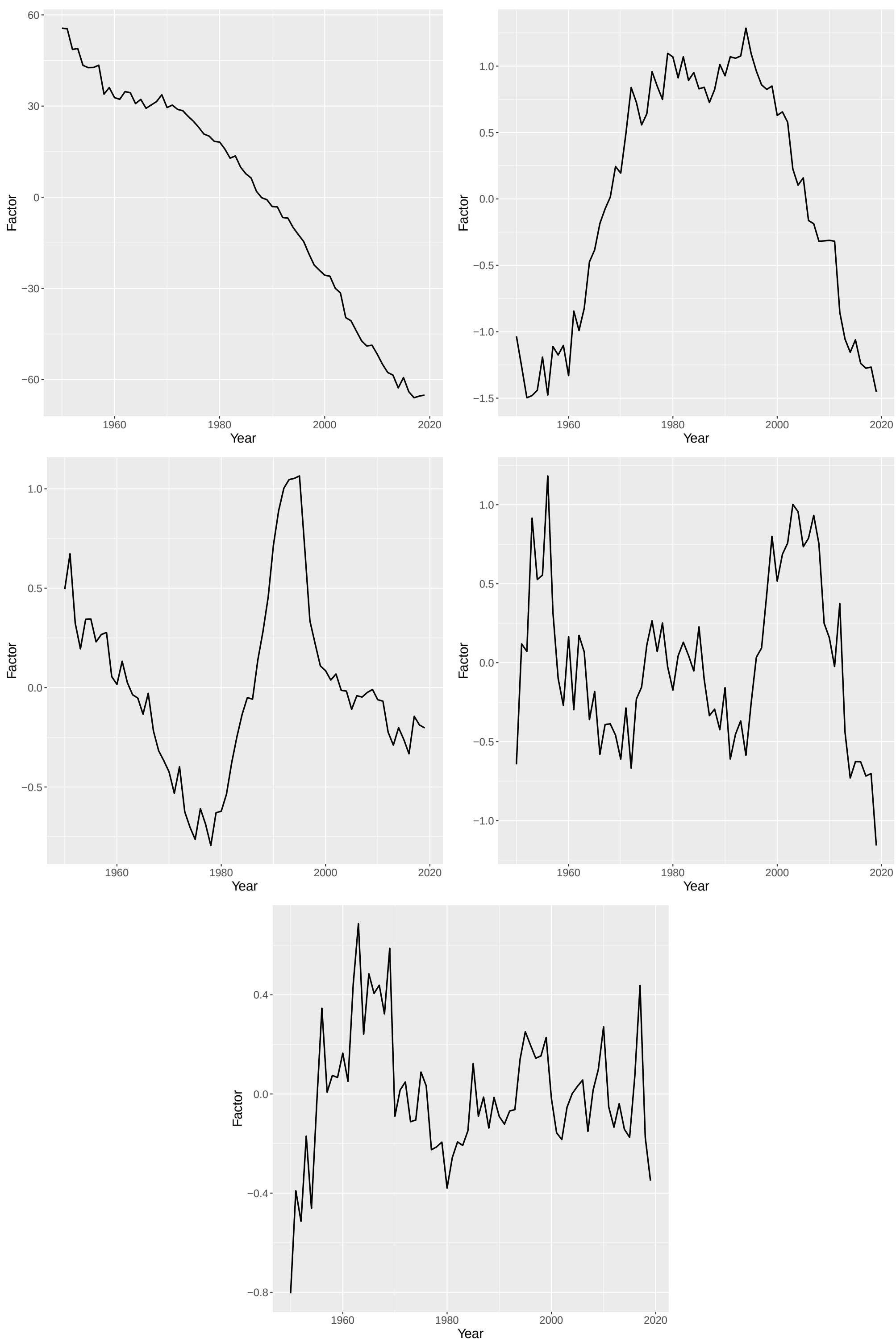

Figure 6: The five latent common factors of the classical multi-factor model.

Figure 6 shows the five fitted latent factors of the classical multi-factor model. (i.e. the Extended LC model) Table 4 shows the correlation matrix of the SDF latent factor and the Extended LC latent factor.

SDF/Extended LC	first latent factor	second latent factor	third latent factor	fourth latent factor	fifth latent factor
first latent factor	0.9448	-0.3193	0.0207	-0.0352	-0.0089
second latent factor	0.2908	-0.0043	0.1153	-0.0582	-0.4534
third latent factor	0.9482	0.0743	0.2821	-0.0191	0.0364
fourth latent factor	0.9900	-0.0227	-0.0536	-0.0476	0.0988
fifth latent factor	0.8073	0.5411	-0.0655	-0.0786	-0.0415

Table 4: Correlation matrix of the SDF latent factor and the Extended LC latent factor

Table 4 shows that there is almost no relationship between both models' latent factors besides the first fitted latent factor of the Extended LC. From Figure 4, we are able to observe that the first fitted latent factor of the Extended LC model is capturing the overall downward dynamic trend of the mortality experience. This might be the reason why it is strongly correlated with a few SDF fitted factors. The second fitted factor captures the impacts of the HIVS/AIDS global pandemic and the rest of the fitted factors are much more complex and it is hard to interpret. Without any further doubt, the SDF model is able to provide factors that are able to be interpreted had shown its superiority where traditional mortality learning methods are not able to provide well interpretable factors.

3.5 Robust Test of the SDF Model

Forecast Horizon	Number of Factor	SDF 1	SDF 2	SDF 3
1	1	**0.01620**	**0.01620**	0.01625
	3	**0.0123**	0.0362	0.0345
	6	**0.0067**	0.0679	0.0413
5	1	**0.01430**	**0.01430**	0.01434
	3	**0.0119**	0.0365	0.0295
	6	**0.0066**	0.0696	0.0375
10	1	0.01240	0.01240	**0.01237**
	3	**0.0112**	0.0393	0.0205
	6	**0.0063**	0.0749	0.0374
15	1	**0.01070**	**0.01070**	0.01073
	3	**0.0101**	0.0376	0.0168
	6	**0.0066**	0.0784	0.0378

Table 5: The in-sample performance when a different set of age factor is used.

We then examine the robustness of the model by investigating the model performance and forecast accuracy when a different set of factors is used. We use two other different sets of age factor along with the set of age factors we selected which is denoted as SDF 1 (i.e. age 0, age 99, age 33, age 21, age 66, and age 18). SDF 2 refers to the second set of factors we select which is age 5, age 96, age 38, age 61, age 22, and age 19. SDF 3 refers to the third set of factors we select which is age 8, age 91, age 41, age 70, age 28 and age 15. Table 5 shows that there is not much change in the in-sample fit when a different set of factors is used.

Forecast Horizon	Number of Factor	SDF 1	SDF 2	SDF 3
1	1	**0.0699**	**0.0699**	0.06992
	2	**0.0093**	0.0109	0.0131
	3	**0.0094**	0.0104	0.0118
	4	**0.0099**	0.01	0.0125
	5	**0.0097**	0.009996	0.0123
	6	**0.0095**	0.0101	0.0123
5	1	**0.0859**	**0.0859**	**0.0859**
	2	**0.0171**	0.0195	0.0184
	3	**0.0169**	0.0184	0.0174
	4	0.0174	**0.0162**	0.0176
	5	0.0167	**0.0166**	0.0178
	6	0.0175	**0.0163**	0.0177
10	1	0.11410	0.11410	**0.11405**
	2	0.0374	0.0411	**0.0370**
	3	0.0365	0.0384	**0.0351**
	4	0.0357	**0.0323**	0.0353
	5	**0.0352**	0.0353	0.0353
	6	0.0364	**0.0345**	0.0357
15	1	**0.14790**	**0.14790**	0.14793
	2	0.0686	0.0704	**0.0647**
	3	0.0692	0.0665	**0.0619**
	4	0.0683	**0.0562**	0.0644
	5	**0.0639**	0.0702	0.0661
	6	0.0678	0.0705	**0.0670**

Table 6: The out-of-sample performance when a different set of age factor is used.

We then list the out-of-sample performance for each sets of factor and it shows that when a different set of factor is used, the forecast accuracy do not fluctuate much as well. Both Table 5 and 6 show that the changes are all very small which suggest that the SDF model is able to produce a better forecast with a high level of robustness.

4 Conclusion

In this paper, the SDF model is introduced as an extension of the multi-factor LC model. It involves a supervised mortality learning method and the factors are selected priory based on bit and piece of expert knowledge of the mortality experience. Two potential forecasting methods, the VAR method and the ARIMA method, had been brought up and after conducting an investigation on the forecast accuracy with respect to the two methods, the ARIMA model provides an outstanding results. A connection between the SDF model and the other three approaches is made by investigating the model performance through MSE across the different forecast horizons. Results were drawn in Figure 3 and it suggests that the SDF model works the best with five factors. In other words, the answers to the research question we stated in Section 1 is that, we are able to improve the model performance with robust mortality forecasts by conducting a supervised mortality learning with factors selected priory as well as by controlling the complexity of the proposed model.

Discussions of the five fitted latent factors as well as the associated latent factors are allowed an ARIMA process had been conducted after the model performance section. Connection of the five fitted SDF factors with the five fitted Extended LC model had been made and it rings the importance of the interpretability that the SDF model is able to provide. We then conclude that studies on mortality forecasts should be considered using a supervised learning method if the mortality data shows a relationship between age and the mortality experience, which is generally the case across the world's population.

Acknowledgments

Zhen Khang Lincoln Sim thanks Professor Heather M. Anderson for feedback on milestones throughout the year, Associate Professor Dan Zhu for supervision and resources throughout the year, and JianJie Shi for coding support.

Declaration of Interest

The authors declare that they have no conflict of interest.

Appendix

Codes used to produce empirical analysis had been included in the R script named Code.R and instructions are given in the Read-me.pdf file. For back up if the files cannot be opened, please refer to https://github.com/ZKLSim/Thesis.git .

References

Bai, J., & Wang, P. (2015). Identification and bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics*, *33*(2), 221–240.

Booth, H., Hyndman, R. J., Tickle, L., & De Jong, P. (2006). Lee-carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic research*, *15*, 289–310.

Booth, H., Maindonald, J., & Smith, L. (2002). Applying lee-carter under conditions of variable mortality decline. *Population studies*, *56*(3), 325–336.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Routledge.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., & Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal*, *13*(1), 1–35.

French, D., & O'Hare, C. (2013). A dynamic factor approach to mortality modeling. *Journal of Forecasting*, *32*(7), 587–599.

Gompertz, B. (1825). Xxiv. on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. frs &c. *Philosophical transactions of the Royal Society of London*(115), 513–583.

Green, P. J., & Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Crc Press.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hyndman, R. J., & Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, *51*(10), 4942–4956.

Kuang, D., Nielsen, B., & Nielsen, J. P. (2008). Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika*, *95*(4), 979–986.

Lazar, D., & Denuit, M. M. (2009). A multivariate time series approach to projected life tables. *Applied Stochastic Models in Business and Industry*, *25*(6), 806–823.

Lee, R., & Miller, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality. *Demography*, *38*(4), 537–549.

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

Li, J., Li, J. S.-H., Tan, C. I., & Tickle, L. (2019). Assessing basis risk in index-based longevity swap transactions. *Annals of Actuarial Science*, *13*(1), 166–197.

Makeham, W. M. (1867). On the law of mortality. *Journal of the Institute of Actuaries (1866-1867)*, *13*(6), 325–358. Retrieved 2022-10-16, from `http://www.jstor.org/stable/41134517`

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115–133.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, *2*(11), 559–572.

Renshaw, A. E., & Haberman, S. (2006). A cohort-based extension to the lee–carter model for mortality reduction factors. *Insurance: Mathematics and economics*, *38*(3), 556–570.

Richman, R., & Wüthrich, M. V. (2021). A neural network extension of the lee–carter model to multiple populations. *Annals of Actuarial Science*, *15*(2), 346–366.

Santosa, F., & Symes, W. W. (1986). Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, *7*(4), 1307-1330. Retrieved from `https://doi.org/10.1137/0907087` doi: 10.1137/0907087

Tang, S., Li, J., & Tickle, L. (2022). A new fourier approach under the lee-carter model for incorporating time-varying age patterns of structural changes. *Risks*, *10*(8), 147.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. Retrieved 2022-09-27, from `http://www.jstor.org/stable/2346178`

Tuljapurkar, S., Li, N., & Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. *Nature*, *405*(6788), 789–792.

Yang, S. S., Yue, J. C., & Huang, H.-C. (2010). Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models. *Insurance: Mathematics and Economics*, *46*(1), 254–270.

Zhu, D., Shi, J., Shi, Y., & WANG, P. (n.d.). Multi-population mortality modelling: A bayesian approach. *Available at SSRN 4127901*.