MASTER'S DEGREE PROJECT IN MATHEMATICS

The efficiency of Hyndman-Ullah methods in case of populations with abnormal short-term increases of mortality rates due to wars and pandemics

by

Teo Raspudić

MAA044 — Examensarbete i matematik för masterexamen

DIVISION OF MATHEMATICS AND PHYSICS
MÄLARDALEN UNIVERSITY
SE-721 23 VÄSTERÅS, SWEDEN

Mälardalen University

MAA044 — Master's Degree Project in Mathematics

Date of presentation:
01 June 2022

Project name:
The efficiency of Hyndman-Ullah methods in case of populations with abnormal short-term increases of mortality rates due to wars and pandemics

Author(s):
Teo Raspudić

Version:
2nd June 2022

Supervisor(s):
Milica Rančić

Examiner:
Anatoliy Malyarenko

Comprising:
30 ECTS credits

Abstract

This report is written with the goal of analyzing the efficiency of the Hyndman-Ullah (HU) and weighted Hyndman-Ullah (wHU) methods when working with the populations that suffered higher mortalities due to the wars and pandemics. Accordingly, the HU and wHU methods are applied to the training sets containing outlying years. The historical data are obtained from Human Mortality Database, while forecasts and analysis are performed using R-package *demography*. Two countries, Sweden and France, are chosen to participate in the in-sample mortality forecast. By comparing forecasted mortalities for those countries after First World War (WWI), Spanish flu and Second World War (WWII), the forecast accuracy of both HU and wHU methods is evaluated. The wHU method proved better when working with the Swedish training set with moderate jumps in mortality. For the French data set consisting of large mortality jumps, both methods recorded significant error measures, which were decreased eventually by extending the training set. After that, the selection of four countries (Sweden, Denmark, Spain and Japan) provides the out-of-sample mortality forecast after the pandemics Covid-19 for the horizon of 30 years.

Keywords: Mortality Forecasting, Hyndman-Ullah, Weighted Hyndman-Ullah, Principal Components Analysis, ARIMA, Life Expectancy, Weighted Penalized Regression Splines, Covid-19.

Acknowledgements

I would like to express my sincere appreciation and gratitude to my supervisor Milica Rancic for the patience, guidance and contribution to the quality of my master thesis, but also throughout my whole study period. I would take this opportunity to highlight the importance of my family throughout this whole process. My deepest thanks are addressed to my mother Vesna Raspudic who endlessly invested herself and provided me with lots of love and support.

Contents

List of Figures

List of Tables

Abbreviations and Acronyms

AIC Akaike Information Criterion

ARIMA Autoregressive Integrated Moving Average

BIC Bayesian Information Criterion

fPCA functional Principal Components Analysis

HMD Human Mortality Database

HU Hyndman-Ullah

h Horizon of the Forecast

i.i.d. independent and identically distributed

LC Lee-Carter

MAE Mean Absolute Error

ME Mean Error

MAPE Mean Absolute Percentage Error

MISE Mean Integrated Squared Error

MPE Mean Percentage Error

MSE Mean Squared Error

PCA Principal Components Analysis

WWI First World War

WWII Second World War

wfPCA weighted functional Principal Components Analysis

wHU weighted Hyndman–Ullah

Chapter 1

Introduction

In today's world, forecasting human mortality occupies a highly important role in the financial long-term plans of the majority of countries. Due to various improvements in di□erent aspects of life, the global life expectancy is higher than ever. Better economic standard, style of life, education, healthcare and quality nutrition are only some of the factors that are rapidly prolonging human lives. Consequently, the number of people in their post productive years is constantly increasing and this trend would significantly a□ect pension funds if not handled appropriately. Therefore, governments need the best estimates of future mortality rates to construct optimal financial strategies. Playing such a notable role in the very structure of the society, the development of this area of actuarial science is of crucial significance.

1.1 Background and Literature Review

The science of mortality modelling has been gradually developing throughout the previous centuries. Back in the day of the epidemic of the bubonic plague of 1665 in London, the pioneer of demography, English statistician John Graunt, published his famous work known as Bills of Mortality. According to [1], it is a collection of the mortality statistics based on the available data on the London population of that time. It represents the first great contribution in this area which has been developing a lot in the upcoming years.

The desire and e□ort for upgrading mortality rate models was the reason for developing the first theoretical approach to mortality rate modelling. In 1725, French mathematician Abraham de Moivre introduced his simple law of mortality expressed as the linear survival function [7]. The next significant step ahead has been made one hundred years later, in 1825 by Benjamin Gompertz. He proposed the law of human mortality [6] that describes mortality depending on age. Gompertz noticed that the human mortality rate follows an exponential increase against time. Hence, the Gompertz' law is expressed in the following way:

$$\mu_x = Bc^x, \quad x > 0, \tag{1.1}$$

where μ_x represents mortality rate at initial age x, and B and c are constants such that $0 < B < 1$, and $c > 1$. However, this model could not properly consider all realistic cases. By [5], for most populations, the force of mortality is not an increasing function of age throughout the whole life span. Because of that, the Gompertz' life tables have been mostly focused on some middle years of the average person's life, while moving very young and old ages out of consideration.

As an improvement to the mentioned model, in 1860, William Makeham extends it by adding a fixed value to the expression in order to upgrade the accuracy of modelling the mortality rate at younger ages. Mathematically, Makeham's law is written as

$$\mu_x = A + Bc^x, \quad x > 0, \tag{1.2}$$

where A is the added constant, independent of age, representing the possibility of fatal accidents, i.e. non-senescent deaths.

Considering the fact that human life evolves under the influence of many factors, researchers had to imitate its nature in order to find the optimal models. This implies the inclusion of stochastic processes that put accurate mortality forecasting on a more advanced level. According to that, mortality models should rather be dynamic instead of static. So, as time passed, more advanced dynamic models have been established.

In modern times, one of the most important ones is Lee–Carter (LC) model [17] introduced in 1992, developed mainly for modelling and forecasting mortality data of the United States. However, its accuracy, together with simplicity, was the reason for its wide usage worldwide. It is a two-factor model with age and time being its main components. Mathematically, the LC model is represented as

$$\ln(m_{x,t}) = a_x + k_t b_x + \varepsilon_{x,t}, \tag{1.3}$$

where $m_{x,t}$ represents central death rate at age x and time t, a_x and b_x are age-specific constants - average log-mortality at x and relative speed of change at x, respectively. Additionally, k_t represents time-varying index of the level of mortality and $\varepsilon_{x,t}$ is the set of random disturbances or simply residual at age x and time t. One of the flaws is that, despite the popularity, the practice has revealed that LC model can not properly handle populations with more non-linear mortality trends. For more information about this fact, look at [24].

Hence, the model development has resumed and numerous extensions have been created to improve the forecasting. Among many, some of the most significant ones are Lee and Miller [16] in 2001, Booth-Maindonald-Smith [3] in 2002, De Jong-Tickle [4] in 2005, Renshaw Haberman [20] in 2006 and Hyndman-Ullah [14] in 2007.

The Hyndman-Ullah (HU) model is of particular interest to this thesis. Initially implemented to Australian fertility data, this method provides accurate forecasting results while successfully handling outliers. According to [14], as an extension to the LC model, HU is based on multiple functional principal components compared to the previous singular principal component approach in LC method. It performs non-parametric smoothing to diminish some of the dataset's intrinsic randomness. However, it is yet not found a way to properly fit all the properties of mortality into one model. In other words, some additional improvements have continued to develop and one such is the weighted Hyndman-Ullah (wHU) method [13] published in 2009. These two models will be more thoroughly discussed in later chapters.

Furthermore, this thesis can be looked at as the extension on some of the research projects made in recent years.

In [30], the author compares the Hyndman–Ullah method to the results obtained from the 1st and 2nd order Lee–Carter models. Additionally, author tests the influence of the Spanish flu (1918 to 1920) on the model and its ability to handle such mortality shock. In the end, mortality rates affected by the Covid-19 pandemic were estimated. The general conclusion is that the HU model performed as superior among those three for the Swedish population data set.

In [19], the author compares the basic HU and the wHU methods, together with the LC method that serves as a benchmark. The two datasets include total populations of the UK and France. The wHU emerged as a winner in this comparison, providing better variation explanation and forecasting abilities.

1.2 Project Description

This thesis includes two main parts. The first part is the theoretical background behind the important mortality forecasting models for this thesis, their strengths and weaknesses and our expectations. The second one is implementation part where, based on the presented knowledge, the models are implemented in R with the help of the package *demography* [9].

The project is further divided into chapters. In Chapter 1, we present the summary of the most important events during the evolution of mortality forecasting.

Chapter 2 introduces and explains the focal points of the topic of mortality forecasting. The methodology of each important concept is described, together with the two methods, Hyndman-Ullah (HU) and weighted Hyndman-Ullah (wHU) that represent the core of the implementation process.

Chapter 3 consists of the practical implementation of the previously explained concepts and models. First, for the selected countries will be made in-sample forecasting to present performances of the HU and wHU methods when applied to the outlying periods. After that, since wHU proved to be slightly more accurate, we will perform an out-of-sample wHU forecast taking into account mortality rates for the selected countries affected by Covid-19.

Chapter 4 summarizes the key points and conclusions made throughout the implementation part. In the end, we will discuss potential future work as a continuation of this project.

The goal of this thesis is to analyze efficiency of the HU and wHU methods for the cases when the test population exhibits abnormal short-term mortality rates. The implementation consists of three main parts that will be explained in more details later in the project in Chapter 3:

- *Training phase* includes selection of the input data and forecasting while the expected output is known.
- *Validation phase* consists of comparing the obtained output with the expected one, evaluating model's precision and efficiency.

- *Testing phase* is performed if the model provides satisfying results in the two previous phases. Based on the real-life data, now with no benchmark to compare the results of the forecast, implement the validated model.

The main causes for an increase in human mortality are wars and diseases, so first we intend to analyse mortality during Spanish flu and two world wars in selected countries. To be more precise, there is applied in-sample forecasting since reliable databases regarding mortality of those periods already exist and we will compare the existing data with our prediction. In that way, it will be shown how efficient and accurate the tested model is. After that training period, we switch to the testing phase. The testing phase is performed with the help of the mortality databases regarding the most recent global pandemic Covid-19. This part belongs to the out-of-sample forecasting because the expected output is not known (pandemic is still ongoing in 2022) and its accuracy will be evaluated when it finally ends.

Chapter 2

Methodology and Model Formulation

In this chapter, the models of the highest interest for this thesis are presented. For complete understanding, first, we will introduce the most important mathematical concepts and then thoroughly explain each model.

2.1 Basic Concepts

The time series forecasting models can be divided into two basic types: univariate and multivariate. Their main difference arises from the number of variables that vary over time. The univariate time series indicates that only one variable is changing so it is classified as a one-dimensional process. It takes only historical values in the process of forecasting. Oppositely, the multivariate time series indicates the change of multiple time-dependent variables and uses external variables to make a prediction.

2.1.1 Autoregressive Integrated Moving Average Model (ARIMA)

Together with exponential smoothing models, ARIMA models are the most popular time series forecasting approaches. With exponential smoothing [10], we tend to explain seasonality and systematic trends where past observations are used to create predictions. On the other side, ARIMA models explain autocorrelations in the data set. The name ARIMA(p, d, q) is the acronym such that:

- $AR(p)$ denotes the autoregressive component [26], meaning that evolving variable is regressed on its previous value. Therefore, p represents the order of the autoregressive process.

- $I(d)$ stands for integrated component [26], meaning that differences among the current and preceding values are replacing data values. Hence, d denotes the number of lag observations and the order of differencing, so the series becomes stationary. The stationary time series is defined to be the time series that is not dependant on observing times. In other words, it looks the same at any point in time [10].

- $MA(q)$ denotes the moving average component [26], meaning that regression error is a linear combination of the forecast errors that occurred in the past. Therefore, q represents the order of the moving average process.

Definition 1. If the linear time series $\{y_t\}$ is an ARIMA(p, d, q), where $p, d, q \in \{\mathbb{Z}^+ \cup \{0\}\}$, then the diﬀerenced series y'_t can be expressed as

$$y'_t = c + \sum_{i=1}^{p} \phi_i y'_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}, \quad t = 1, ..., N, \tag{2.1}$$

where c is the constant value, $\phi_1, ..., \phi_p$ is the set of $AR(p)$ model parameters, $\theta_1, ..., \theta_q$ is the set of $MA(q)$ model parameters and ε_t denotes the error term.

The diﬀerenced series can be defined as the series of changes between all consecutive observations from the initial series, mathematically written as

$$y'_t = y_t - y_{t-1}, \quad t = 1, ..., N, . \tag{2.2}$$

Because it is not possible to find the change for the first observation (the 0th observation does not exist), the diﬀerenced series contains in total $N - 1$ values.

The model ARIMA$(1, 1, 1)$ indicates that the time series is stationary at the first diﬀerence while the underlying AR and MA models are of the first-order. On the same principles, there exist several special cases of ARIMA models and they are [10]:

- ARIMA$(p, 0, 0)$ or simply AR(p) represents generalization of the autoregressive process of the order p that can be expressed as

  $$r_t = c + \phi_1 r_{t-1} + ... + \phi_p r_{t-p} + \varepsilon_t,$$

  where c is the constant term and $\phi_1, ..., \phi_p$ is the set of model parameters. Also, r_t and r_{t-1} are dependant and explanatory variables, respectively, and ε_t denotes a white noise series with mean zero and variance σ^2 [26].

- ARIMA$(0, 0, q)$ or simply MA(q) represents generalization of the moving average process of the order q that can be expressed as

  $$r_t = c + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q},$$

  where c is a constant, θ_q denotes the model parameter and ε_t represents a white noise.

- ARIMA$(0, 0, 0)$ represents the *white noise* that is defined as a stationary time series in which all data are i.i.d. with the mean zero and variance σ^2, implying that there is no correlation between any pair of values.

- ARIMA$(0,1,0)$ without a constant represents *random walk* which is primarily used for non-stationary data. The model is written in the following way:

$$y_t = y_{t-1} + \varepsilon_t,$$

where y_0 is the first point of the series and ε_t is a zero-mean white noise with constant finite variance σ_ε^2. Hence, the true random walk is distributed as $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$. From the model expression, we could notice that random walk bases its forecast exclusively on the last observation. Due to the lack of additional information, it is not possible to predict future movements.

- ARIMA$(0,1,0)$ with a constant represents *random walk with drift* that is written as

$$y_t = c + y_{t-1} + \varepsilon_t,$$

where c is added constant term expressed as $E\left[y_t - y_{t-1}\right]$.

The particular importance of ARIMA models is the fact that, as authors in [15] explain, they are more suitable representations of mortality time series than simple random-walk models. In the implementation part of this thesis, Chapter 3, the order of the concrete ARIMA model will be determined with the help of the R package *demography*.

2.1.2 Functional Principal Component Analysis (fPCA)

The principal component analysis (PCA) is a widely used method that tackles problems of reducing the dimensionality of the large data sets while minimizing the loss of information. It is a measure of variability, where each principal component possesses information of the part of the total data set's variance. However, the PCA is a *discrete concept* and its limitations in the analysis of data that are interpreted as a curve or function, instead of a discrete sequence of observations, are visible. Therefore, we will use the fPCA that is defined as a technique of dimension reduction for multivariate data, extended to *time series*. The role of the fPCA is to ease obtaining of the finite-dimensional vector of random scores by transforming inherently infinite-dimensional functional data [29]. This inheritance appears due to infinitely many points in time in the continuous data set.

Since in the HU method the mortality is a continuous and smooth function of age $(f_t(x))$, the regular PCA for functions of discrete values is not helpful anymore. For that reason, it is essential to use the fPCA to decompose a set of curves into two parts - orthogonal functional principal components and their uncorrelated principal component scores.

The fPCA expansion is expressed as

$$X_t(x) = \mu(x) + \sum_{k=1}^{\infty} \beta_{t,k} \phi_k(x), \tag{2.3}$$

where

$$\beta_{t,k} = \int_I (X_t(x) - \mu(x))\, \phi_k(x) dt,$$

stands for the scores of the random process X_t and $\phi_k(x)$ is the set of orthogonal functional principal components. The $\beta_{t,k}$ regulates the dynamics of the process and represents the uncorrelated principal component scores, which are independent across t for a sample of independent trajectories and are uncorrelated across k. The expected value and variance are then $E(\beta_{t,k}) = 0$ and $Var(\beta_{t,k}) = \lambda_k$, respectively, where λ_k are the eigenvalues in descending order [29]. Furthermore, for large enough K (see 2.2.1), the equation (2.3) is a fine approximation to the infinite sum which can then be written as

$$X_{t,K}(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x). \tag{2.4}$$

In this way, information carried in X_t is successfully transferred into K-dimensional vector $\boldsymbol{\beta}_t = \{\beta_{t,1}, ..., \beta_{t,K}\}$.

In [14], Hyndman and Ullah propose the two-step algorithm for functional principal components:

Step 1. Apply the RAPCA (Robust Adaptive Principal Component Analysis) algorithm to find the initial and very robust values of

$$\{\beta_{1,k}, ..., \beta_{N,K}\} \text{ and } \{\phi_1(x), ..., \phi_K(x)\}, \quad k = 1, ..., K.$$

The RAPCA algorithm has been first introduced by [8] and is characterized by numerical stability and time e□ciency. It can work only with discrete multivariate data and consequently, we must perform discretization of the continuous smooth function $f_t(x)$. In that way, the resulting values are approximated. For more information about the RAPCA algorithm, see [14].

Step 2. Express the integrated squared error $v_t(x)$ for year t in the following way:

$$v_t(x) = \int_x \left(\hat{f}_t^*(x) \sum_{k=1}^{K} \hat{\beta}_{t,k}\phi_k(x) \right)^2 dx$$

in order to obtain the measure of the accuracy of the principal component approximation for that year t. If median of the set $\{v_t\}$ is denoted by s and the controlling parameter of the robustness degree represented by λ, then the weights are:

$$w_t = \begin{cases} 1, & \text{if } v_t < s + \lambda\sqrt{s}, \quad \lambda > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

With those computed weights, find the new weighted estimates of $\{\hat{\beta}_{t,k}\}$ and $\{\phi_k(x)\}$ by applying the exact procedure thoroughly explained in [14].

As it may be noticed, this approach considers binary rather than continuous weights. The main reason is the correlation of demographic data and its outliers that are most commonly caused by two factors, wars and pandemics. Since both of these outlying periods generally last relatively short, the usual practice is to remove outlying years but to keep all other available information in the forecasting process. Because of that, there is no need to employ continuous weights, $w_t = v_t^{-1}$, and assigning weights to equal 0 or 1 is justified for practical reasons.

2.1.3 Weighted Penalized Regression Splines

In [14], Hyndman and Ullah emphasize the importance of applying weighted penalized regression splines with a partial monotonic constraint in the process of smoothing.

Definition 2. Let $m_t(x)$ symbolize the observed mortality rate at age x in year t. Also, let $N_t(x)$ represent the total population of age x on particular date in year t. Then, it is assumed that $m_t(x)$ is binomially distributed with the variance

$$Var[m_t(x)] = \frac{[1 - m_t(x)]\, m_t(x)}{N_t(x)}.$$

Therefore, by a Taylor approximation, the variance of $y_t(x) = \log(m_t(x))$ can be expressed as

$$\hat{\sigma}_t^2(x) = \frac{1 - m_t(x)}{N_t(x) m_t(x)}. \tag{2.6}$$

The estimation of the curve $f_t(x)$ for each year is performed thanks to the weighted penalized regression splines. Since the weights, denoted by $w_t(x)$ are defined to be the inverse variances, they can be expressed as

$$w_t(x) = \frac{N_t(x) m_t(x)}{1 - m_t(x)}. \tag{2.7}$$

As already mentioned, in the HU method presented in the succeeding section, weighted penalized regression splines with a partial monotonic constraint are used. The purpose of that partial monotonic constraint is to reduce the noisy log mortality rates at older ages and generally decrease the overall variability. For example, at sufficiently old age such as 50 years, it is assumed that mortality is continuously increasing for all upcoming years.

2.1.4 Life Expectancy

The available and trustworthy data from recent centuries suggest that most countries worldwide observe that future generations have longer average lifespans than their ancestors. In other words, life expectancy has increased in the majority of countries across the world. However, this tendency is sometimes disrupted for different reasons. According to [5], the three main causes of the mortality change over time are: trends, shocks and idiosyncratic causes. To a□rm the existence of mortality trends, it is usually su□cient to recognize mortality patterns throughout a certain period. The trend represents the moderate change in mortality across years. Contrary, mortality shock is the sudden increase in mortality rates caused by war or epidemics. Due to its unpredictable nature and relatively short-lasting duration, the mortality shock can not be classified as a pattern. Another unpredictable cause is an idiosyncratic cause characterized by the random variation not induced by the trend or shock. For instance, the development of the new drug that unpredictably negatively a□ects people of the certain age or health state, thus causing temporary jump in the mortality until it is not removed from the market.

The mortality trends are in close relation to life expectancy. The life expectancy represents the average number of years that the individual can expect to live according to the age-specific death rates. As the authors in [2] conclude, the relation between the relative accuracy of the forecasted life expectancy and modelled log death rates is not strong. In Chapter 3, together with the in-sample forecasting of mortality rates of selected countries, we will forecast life expectancy across all ages in the given forecasting horizon and evaluate the relationship between those measures.

2.2 The Hyndman–Ullah (HU) Method

The HU method has been introduced as an answer to the limitations of the LC model, primarily the di□culties of facing non-linear mortality trends. In 2007, Hyndman and Ullah constructed the new, improved method for modelling and forecasting age-specific mortality and fertility rates [14], as well as log mortality rates [22]. The main characteristics of the Hyndman–Ullah method can be summarized in the following way:

(1) HU method starts with the assumption of mortality being the continuous and smooth function of age, $f_t(x)$, that has an observed error at discrete ages. Before modelling, a non-parametric smoothing of the log mortality rates using weighted penalized regression splines with a partial monotonic constraint is performed. Thus, we can look at age x as the continuous variable of the function $m_t(x)$.

Definition 3. Let $y_t(x)$ be the logarithm of the observed death rate $m_t(x)$ for age $x \in [x_1, x_p]$ in year t. Then, it stands that

$$y_t(x_i) = \log[m_t(x)] = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}, \quad (2.8)$$

for

$$t = 1, \dots, N \quad \text{and} \quad i = 1, \dots, p,$$

where

- $f_t(x_i)$ stands for the smooth function of age x in year t,
- $\sigma_t(x_i)$ denotes the noise component that varies with x_i in year t,

- $\varepsilon_{t,i}$ is an i.i.d. standard normal random variable $(\varepsilon_{t,i} \sim \mathcal{N}(0, \sigma^2))$, indicating that the expected value is zero and σ^2 is the variance.

According to [14], because x is the continuous variable, in most of the cases $\{x_1, ..., x_p\}$ represents either single years of age or age groups in the range of five years.

(2) Compared to the LC model [17] that tends to capture the mortality rate patterns using singular principal component, the HU model uses principal components of higher-order to capture additional dimensions of change in mortality rates [23]. In that way, multiple principal components may capture the remaining non-random patterns that went unrecognized by only one principal component.

Definition 4. Let the set of curves be decomposed into orthogonal functional principal components $\{\phi_1(x), ..., \phi_K(x)\}$ and their uncorrelated principal component scores $\{\beta_{t,1}, ..., \beta_{t,K}\}$. Therefore, using fPCA, we obtain $f_t(x)$ such that

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x), \tag{2.9}$$

where

- $\mu(x)$ is the mean function across ages estimated by

$$\hat{\mu}(N) = \frac{1}{N}\sum_{t=1}^{N} f_t(x),$$

- K denotes the number of principal components while $K < N$,
- $\beta_{t,k}$ stands for a set of uncorrelated principal component scores and it is the time series coe□cient function that corresponds to the k_t in LC method [23],
- $\phi_k(x)$ represents set of the K orthogonal functional principal components,
- $e_t(x)$ is the residual function with mean zero.

In Definition 3, $\sigma_t(x_i)\varepsilon_{t,i}$ represents the di□erence between the observed rates and the spline curves. In other words, it represents the error of observation that changes with age.

In Definition 4, the error term $e_t(x)$ represents the di□erence between the fitted curves from the model and the spline curves. Hence, it can be called the modelling error [2].

(3) A random walk with drift model implemented in [17] is not su□cient anymore but instead, the forecasting of the principal component scores can be made using various univariate time series models. The best examples are the autoregressive integrated moving average (ARIMA) model or, as the authors presented in [12] and [14], the exponential smoothing state space models.

(4) Into the nature of HU method are incorporated concepts of functional principal component regression, functional data analysis, non-parametric smoothing and robust statistics [14]. As a result, it significantly better manipulates outliers in the data set. This property will play an important role in the implementation part of this thesis since we are forecasting mortality rates after the two world wars and pandemics of Spanish flu and Covid-19 in selected countries.

(5) The constant mortality index k_t is not adjusted as it is in the LC method. For more details about adjusting k_t, see [2]. The HU model is flexible enough to take notable covariates, for example, medical treatments, screening tests, etc. into consideration in the process of modelling, providing more accurate predictions.

2.2.1 Forecasting

As authors of [11] explain, the influence of the di□erent numbers of principal components K has been investigated throughout the years. The small K highly decreases the model accuracy due to the increase in bias and consequently, the increase in approximation error. On the other side, too large K prolongs computation time and intensifies variance because more components need to be estimated, adding random error. For large enough K, the forecasting results do not di□er so there is no need for selecting some especially large values.

Hyndman and Ullah's decision had to be based on minimizing the mean integrated squared forecast error, or shortly MISE, defined as:

$$\text{MISE} = \frac{1}{N}\sum_{t=1}^{N} e_t^2(x)dx.$$

Considering that, they found that for their data set, the optimal decision would be $K = 6$ principal components. According to the constraint $K < N$, it can be concluded that 7 or more time periods should be forecasted in that case. However, the optimal number of principal components highly depends on the data set, so in the implementation part of this thesis we will find out how many we need to use. HU method selects the optimal model based on the most commonly used penalized model selection criterions such are Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Since fPCA is the foundation of the HU method, Hyndman and Ullah in [14] forecast independently each univariate time series $\{\hat{\beta}_{t,k}\}$ for $k = 1, .., K$. Finally, to estimate future curves, they find the products of forecasted principal component scores and the principal components.

The fusion of equations (2.8) and (2.9) provides the compact form of the single equation that is written as

$$y_t(x_i) = \mu(x_i) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x_i) + e_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}. \tag{2.10}$$

The authors in [14] performed conditioning on the set of all observed data denoted by

$$\mathcal{I} = \{y_t(x_i); \;\; t = 1, ..., N; \;\; i = 1, ..., p\},$$

and the set of functional principal components $\mathbf{\Phi}$, so they obtained h-step ahead forecasts of $\hat{y}_{N+h}(x)$ expressed as

$$\hat{y}_{N,h}(x) = E[y_{N+h}(x)|\mathcal{I}, \mathbf{\Phi}] = \hat{\mu}(x) + \sum_{k=1}^{K} \tilde{\beta}_{N,k,h}\hat{\phi}_k(x), \tag{2.11}$$

where $\tilde{\beta}_{N,k,h}$ represents h-step ahead forecast of $\beta_{N+h,k}$ using a univariate time series model. Furthermore, from the combined equation (2.10), the forecast variance can be presented as

$$\zeta_{N,h}(x) = Var[y_{N+h}(x)|\mathcal{I}, \mathbf{\Phi}] = \hat{\sigma}^2_{\mu}(x) + \sum_{k=1}^{K} u_{N+h,k}\hat{\phi}^2_k(x) + v(x) + \sigma^2_t(x), \tag{2.12}$$

where $u_{N+h,k} = Var(\beta_{N+h,k}|\beta_{1,k}, ..., \beta_{N,k})$ is the smoothing error found from the time series model, $\hat{\sigma}^2_{\mu}(x)$ is the variance of the smooth estimate $\hat{\mu}(x)$, representing an error of predicting the dynamics, and $v(t)$ denotes the model error (residual) variance that is estimated by computing the average of $\hat{e}^2_t(x) = \{e^2_1(x), ..., e^2_N(x)\}$ for each age x.

Since principal components and the error term are orthogonal in the HU method, the approximate overall forecast variance can be calculated as a sum of all component variances. Assuming non-correlation and normal distribution of the di□erent types of errors, the prediction interval for $y_{N+h}(x)$ has the form

$$\hat{y}_{N,h}(x) \pm 1.96\sqrt{\zeta_{N,h}(x)}.$$

2.2.2 Weighted Functional Principal Component Analysis (wfPCA)

This subtopic can be looked at as the introduction to the extended version of the HU method known as the weighted Hyndman-Ullah method (wHU). The new approach involves geometrically receding weights in the principal component decomposition, which essentially means that modern demographic data weigh more than the older ones. In the wfPCA, the estimation of the mean function μ_x is done by employing a weighted average (see Definition 4) in the following way [13]:

$$\hat{\mu}(x) = \sum_{t=1}^{n} w_t \hat{f}_t(x), \tag{2.13}$$

where $\hat{f}_t(x)$ represents a smoothed curve obtained from $y_t(x)$. The weights w_t are expressed as

$$w_t = \kappa(1 - \kappa)^{n-t},$$

where $0 < \kappa < 1$ and κ denotes the optimal value found through the minimization of the overall forecast error. Then, the fPCA explained in Subsection 2.1.2 can be implemented to the mean-or median-adjusted functional data $\{\hat{f}_t^*(x)\}$ with the aim of obtaining principal components $\{\phi_k(x)\}$ and their scores $\{\beta_{t,k}\}$ [13]. Here, $\hat{f}_t^*(x)$ is defined as

$$\hat{f}_t^*(x) = \hat{f}_t(x) - \hat{\mu}(x).$$

Based on the mentioned, the new wHU model is presented and explained in the following section.

2.3 The Weighted Hyndman–Ullah Method (wHU)

Due to many di erent aspects of civilization's progress, life conditions are constantly changing during time. Substantial improvements in hygiene, nutrition, medicine and overall lifestyle positively a ected the global average life expectancy and decreased the mortality rates. Accordingly, the more optimized mortality forecasting models expressed the need for the unequal perception of the historical data, depending on how old the observations are. The role of the newer observations must be prioritised because of their similarities and relevance to the actual trends. This idea lead to the establishment of the improved version of the HU method that includes weights. As already mentioned in the wfPCA subsection, the wHU method allows greater weight to be assigned to more recent observations than those from the long past.

Definition 5. Let the set of weighted curves be decomposed into orthogonal functional principal components $\{\phi_1(x), ..., \phi_K(x)\}$ and their uncorrelated principal component scores $\{\beta_{t,1}, ..., \beta_{t,K}\}$ [23]. Therefore, using fPCA, we obtain $f_t(x)$ such that

$$f_t(x) = \hat{\mu}(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x), \tag{2.14}$$

where $\hat{\mu}(x)$ is the weighted functional mean function across ages estimated by the weighted average from equaltion (2.13).

2.3.1 Forecasting

Similarly as it was in the HU method, the fusion of equations (2.8) and (2.14) o ers the compact form of the single equation, written as

$$y_t(x_i) = \hat{\mu}(x_i) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x_i) + \hat{e}_t(x_i) + \hat{\sigma}_t(x_i)\hat{\varepsilon}_{t,i}. \tag{2.15}$$

Further, the authors in [13] performed conditioning on the set of all observed data denoted by

$$\mathcal{I} = \{y_t(x_i);\ \ t = 1, ..., N;\ \ i = 1, ..., p\},$$

and the set of functional principal components $\mathbf{\Phi}$, resulting in getting the h-step ahead forecasts of $\hat{y}_{N+h}(x)$ denoted as

$$\hat{y}_{N,h}(x) = E[y_{N+h}(x)|\mathcal{I}, \mathbf{\Phi}] = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{N,k,h}\hat{\phi}_k(x), \tag{2.16}$$

where $\hat{\beta}_{N,k,h}$ represents h-step ahead forecast of $\beta_{N+h,k}$ using a univariate time series model. As it could be noticed, HU and wHU models are quite similar and the only distinction between them is the allocation of weights. Instead of strictly taking $w_t = 0$ in case of outlier and $w_t = 1$ otherwise, wHU model induces weights that decay geometrically to assign more relevance to the later observations [13].

2.3.2 Error Measures

Besides graphical representations of the forecasts, in-sample forecasting provides the numerical representation of the model accuracy. By comparing the obtained results with the actual historical data, it is possible to calculate various types of error measures, which are a helpful indication of the quality of the forecasts. In the implementation part, following the package demography [9], five errors averaged across time and ages will be calculated and they are:

- Mean error (ME) represents the measure of bias. It is defined as the average value of all observed positive and negative di□erences between actual values, denoted by A_t, and measured (forecasted) values, denoted by F_t, in a given set. Accordingly, it can be calculated as

$$\text{ME} = \frac{1}{N}\sum_{t=1}^{N}(A_t - F_t). \tag{2.17}$$

  The main flaw of the mean error measure reflects when the individual errors have both signs (positive and negative) since they cancel each other and give a much smaller average error.

- Mean absolute error (MAE), often called mean absolute deviation (MAD), belongs to the scale-dependent error metrics. The mathematical formula is

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}|A_t - F_t|. \tag{2.18}$$

  Being based on the absolute value function, MAE measures accuracy regardless of sign.

- Mean squared error (MSE) is the scale-dependent error that measures the variation of forecast errors. It can be expressed as

$$\text{MSE} = \frac{1}{N}\sum_{t=1}^{N}(A_t - F_t)^2. \tag{2.19}$$

  It is similar to MAE as a measure of variation, but it is based on the squared function. Because of that, it is characterized by sensitivity to errors that are less than one.

- Mean percentage error (MPE) is determined by finding the mean of the sum of the percentage errors. Accordingly, it can be calculated as

$$\text{MPE} = \frac{1}{N}\sum_{t=1}^{N}\frac{A_t - F_t}{A_t}\cdot 100\%. \tag{2.20}$$

  MPE belongs to the percentage error metrics. Since it represents the size of the error in the form of a percentage, it is identified as the scale-independent measure.

- Mean absolute percentage error (MAPE) follows the same procedure as MPE but takes into consideration the absolute values of each particular forecast error. Therefore, the formula is

$$\text{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\frac{|A_t - F_t|}{A_t}\cdot 100\%. \tag{2.21}$$

  Due to the similarities to MPE, MAPE is also identified as the scale-independent measure.

To find more information about the forecast measures, see [21] and [28] .

Chapter 3

Implementation

The topic of this thesis is in the close connection to the previously analyzed problems described in theses [30] and [19]. In the former one, the author forecasts mortality rates and compares the 1st and 2nd order Lee–Carter methods with the robust HU method. In the latter, the author is forecasting mortality rates by applying the HU and the wHU methods and makes the comparison to the LC method, which serves as the benchmark. They both reached a similar conclusion; the HU method performed better than the LC model, while the wHU method was superior to others for the respective analyzed problem. The task of this report can be viewed as the partial continuation of both of the mentioned theses. In [30], after a thorough presentation of death rates forecasts, the author ends the report by making a crude estimation of mortalities from 2019 to 2022, which is taken as the estimated duration of the Covid-19 pandemic at the moment of writing. The HU method is applied in the process of forecasting and it provides very satisfying results. However, a certain room for error has to be left since the forecast was made based on a hypothetical situation. On the other side, notable efficiency of the wHU method has been presented in [19] while forecasting mortality rates until 2050 for the selected countries. However, it is important to mention that the global Covid-19 mortality increase has not been considered since the data sample included the years up to 2018. Therefore, in this thesis, we are interested in analyzing the efficiency of the HU and wHU models when facing outliers in the mortality data, with the main focus of forecasting mortality during and after the Covid-19 pandemic.

3.1 Historical Data Sets

Throughout the whole thesis, the Human Mortality Database (HMD) [27] is the main source of reliable historical data regarding mortality rates. The selected countries for the in-sample forecasting part are Sweden and France. Knowing that populations of those countries have not been equally affected in the periods of the two world wars, in those cases, we expect different conclusions for each country. Particularly, France's mortality is expected to show more outlying years due to its active role in the First and the Second World Wars (WWI and WWII), while Sweden primarily held the neutral position in both of them. In the period of the Spanish flu, we might expect much closer mortality rates because diseases should not have correlations to politics in the way the wars do. However, the fact that can not be forgotten is that the spread of Spanish flu escalated in the end and just after the First World War, which makes it harder to distinguish the exact cause for many deaths. Therefore, it is reasonable to assume that France recorded at least slightly higher mortality rates during the Spanish flu as well.

To explicitly show what outlying years of mortality rates look like, we will observe the historical data collected for the whole 20th century. In the Figs. 3.1 and 3.2 presented below, we can observe almost a constant decrease in the log mortality rates as time goes by. For example, compare years from the beginning and the end of the 20th century, coloured in red and purple, respectively. We can notice that log mortality rates are higher for all age groups in the first years of the previous century and this is true for both Sweden and France without exceptions. It is proof of the improvement in various living conditions and medical advances throughout the years. Figure 3.1 represents the mortality rates for Sweden in the period between the years 1901 to 2000.

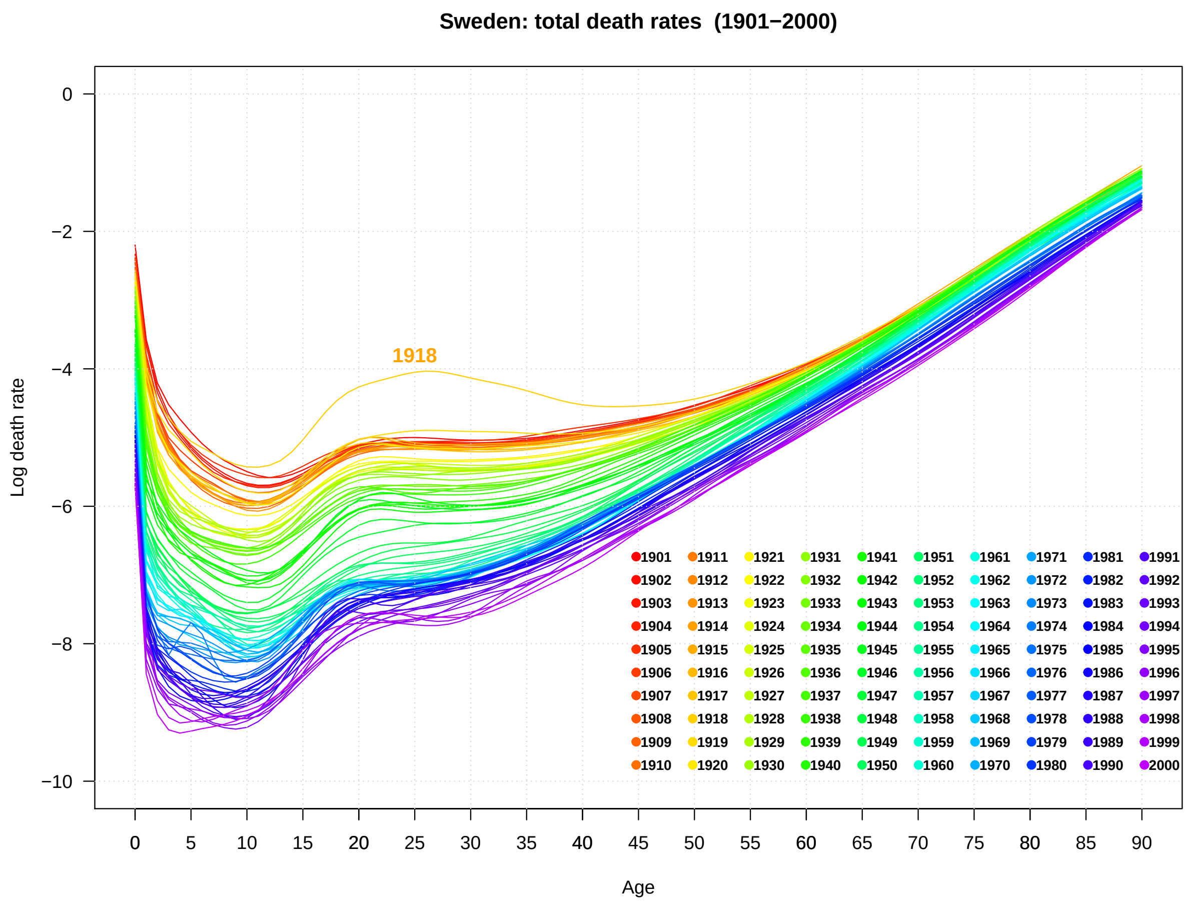

Figure 3.1: The Swedish mortality rates throughout the 20th century.

According to the graph above, Sweden recorded a certain increase in mortality rates in the second decade of the 20th century, but the only obvious outlying year was 1918, coloured in one of the shades of orange. It is the intersection point of the end of the WWI and the global expansion of the Spanish flu. In war times, the mortality of the part of the population that is fit for military service rapidly increases. On the other side, various age groups have been strongly hit during the Spanish flu, especially young adults [31].

It is visible that the whole childhood period and up until the age of approximately 40 years of age had an unusually high mortality rate. The combination of those two major events certainly left a mark on the country's population. However, as was expected, Sweden's mainly neutral position in the critical years of international conflicts significantly contributed to the preservation of lower mortality rates.

Figure 3.2 represents the mortality rates for France in the period between the years 1901 to 2000.

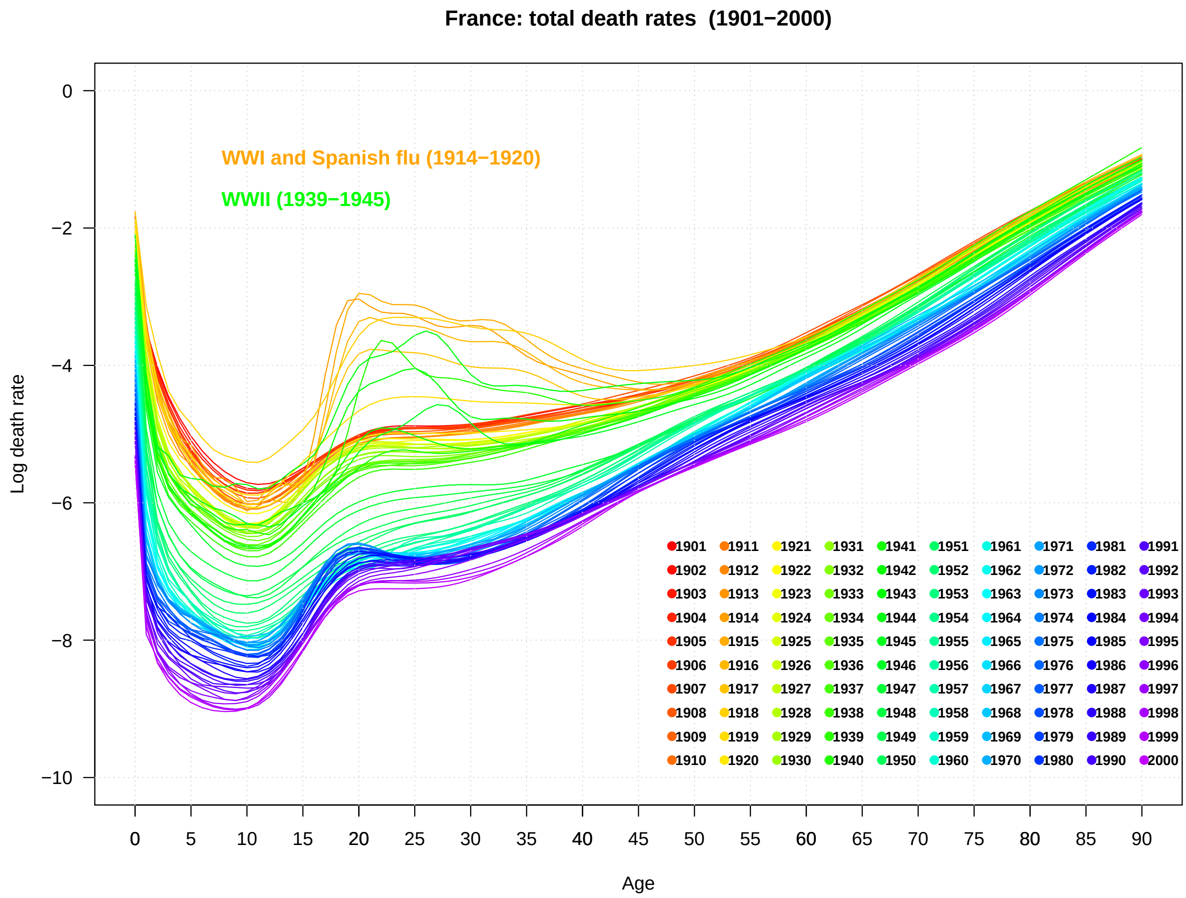

Figure 3.2: The French mortality rates throughout the 20th century.

Let us take a look at the obtained graph. The specific periods of the 20th century seem to be clear outliers from the relatively steady trend of the death rate fall. The critical outlying years are coloured in different shades of orange and green. If we identify the years of the WWI (1914 – 1918), Spanish flu (1918 – 1920) and WWII (1939 – 1945), we may notice that population between approximately 15 and 50 years of age experienced a significant increase in the mortality rate. It is entirely reasonable since this age group fulfilled the military service and, consequently, suffered the most.

3.2 In-Sample Forecasting

The process of in-sample forecasting consists of two major phases - the training phase and the validation phase. In the following pages, we will present the practical implementation of all steps of the in-sample forecast. That will include forecasting mortality rates using HU and wHU methods and calculating different types of forecasting errors. In addition, due to the close connection between mortality and life expectancy, we will construct the table that compares the historical and forecasted life expectancies for the selected countries, Sweden and France.

The purpose of the training phase is to perform the forecast with the already known output. Then, in the validation phase, we seize the opportunity to compare the actual data with the results of the forecast. In that way, we get a quality insight into the performances of the implemented forecasting models when working with the chosen data set. Therefore, from the historical data set of both selected countries, Sweden and France, we extract an arbitrary, but reasonably long period of years. This period is further divided into two parts, the training set that consists of the majority of data and the smaller, testing set that presents forecasting results.

Since it is hard to differentiate the impacts of the First World War from the pandemic of Spanish flu in the end and after the war, the two outlying events will not be separated in our analysis. To forecast mortality rates during the years of WWI and the Spanish flu, the selected period spreads from 1872 to 1925, including those years. By following a similar pattern, the selected period for the forecast of WWII lasts from 1872 to 1950, including those years. However, since this period includes the WWI and Spanish flu which are outlying years, one additional forecast will be performed. We are interested in potential differences in the forecast when its training set does not include the years before, for example, 1924, i.e. excluding WWI and Spanish flu.

The purpose of the validation phase is to check how efficient the applied HU and wHU methods are by comparing the calculated results and the actual data. With the help of the built-in functions in the package demography [9], we will compute five errors averaged across time and ages - mean error (ME), mean absolute error (MAE), mean squared error (MSE), mean percentage error (MPE) and mean absolute percentage error (MAPE).

Finally, due to the close connection between life expectancy and mortality rate, we will construct the table of life expectancies for both Sweden and France across all ages during the forecasting period. In other words, we are interested in observing how HU and wHU methods handle outlying periods when forecasting life expectancy. Afterwards, we will draw a parallel to the mortality rate forecasts and obtain conclusions about their relationship.

3.2.1 WWI and Spanish Flu

The first outlying period that is going to be tested is the period of the First World War, together with the Spanish flu. Therefore, the critical years are in the range between 1914 to 1920. The intention is to forecast the mortality rates in the years after that critical period. Concretely, let the training set cover the years from 1872 until 1925, including them. The reason why this period is not longer is due to the Paris Commune that took place in 1871 [25]. In that way, we try to avoid influencing the forecast by the additional outlying year. That being said, let the testing set covers the interval of the following 13 years, that is from 1926 until 1938, including those years. The reason why we stop at 1938 is that WWII started in 1939 which belongs to another outlying period and neither method has that information so it would not be of any use to forecast anything further. Since the total range of years is 67, the training set consists of 80.6% of the total data, while the testing set includes the remaining 19.4%.

According to the Section 2.2.1, the optimal number of principal components di�ers depending on the data set. Our main goal is to maximize explained variation by increasing the number of principal components. However, at some point, further increase will not result in significantly enlarged explained variation, indicating that the optimal number is reached. Throughout this whole report, neither percentage variation explanation will drop below 99%.

The HU method - For the selected Swedish data set, the optimal number of principal components is $K = 5$. Accordingly, the explained percentage variation due to basis functions is 83.82%, 10.84%, 3.56%, 0.66% and 0.40%, respectively. That gives the sum of 99.28%, while only 0.72 of the variation is left unexplained.

For the selected French data set, the optimal number of principal components is $K = 7$. Therefore, the basis functions explain 67.93%, 27.12%, 1.91%, 1.08%, 0.65%, 0.34% and 0.32%, respectively. That sums to the total of 99.35%, while only 0.65 of the variation is left unexplained.

The wHU method - Considering the selected Swedish data set for $K = 5$ principal components, the explained percentage variation due to basis functions is 92.06%, 5.32%, 1.73%, 0.32% and 0.21%, respectively. That means that the total is 99.64%, while only 0.36 of the variation is left unexplained.

Considering the selected French data set for $K = 7$ principal components, the basis functions explain 54.73%, 41.57%, 1.45%, 0.80%, 0.48%, 0.25% and 0.24%, respectively. That sums to the total of 99.52%, while only 0.48% of the variation is left unexplained.

For the same number of principal components, the wHU method explains more percentage variation for both countries compared to the HU method. Also, no significant di�erence between the two methods has been noticed in the plots. The wHU method is superior in the case when there are not many outlying years, while the HU method returned a better forecast when working with a very volatile data set.

In the Table 3.1 below, numerical values of each of the explained error measures of the forecasted mortality for the First World War and Spanish flu are presented.

Table 3.1: Mean errors of the in-sample forecast of the WWI and Spanish flu for Sweden and France based on the training set $(1872-1925)$.

	Sweden					France				
	ME	MAE	MSE	MPE	MAPE	ME	MAE	MSE	MPE	MAPE
HU	0.0841	0.1543	0.0432	0.1086	0.1755	0.2473	0.2490	0.1113	0.3158	0.3175
wHU	0.0712	0.1151	0.0281	0.0868	0.1292	0.2619	0.2647	0.1278	0.3415	0.3443

Based on the calculations provided in the table above, di erent conclusions can be taken depending on the country. Since Swedish history in the tested period was not nearly as turbulent as French, both methods make forecasts in accordance with the expectation. The wHU method performs better for the moderately outlying period such as Swedish. On the other side, handling the data set with a highly turbulent history such as French caused both methods to return larger errors. However, the HU method performs with better accuracy for this French data set.

Figure 3.3 below provides the comparison of the Swedish historical log mortality rates and the mortality rates forecasted by HU and wHU methods. On all graphs, the x-axis represents age and the y-axis represents log mortality rate.

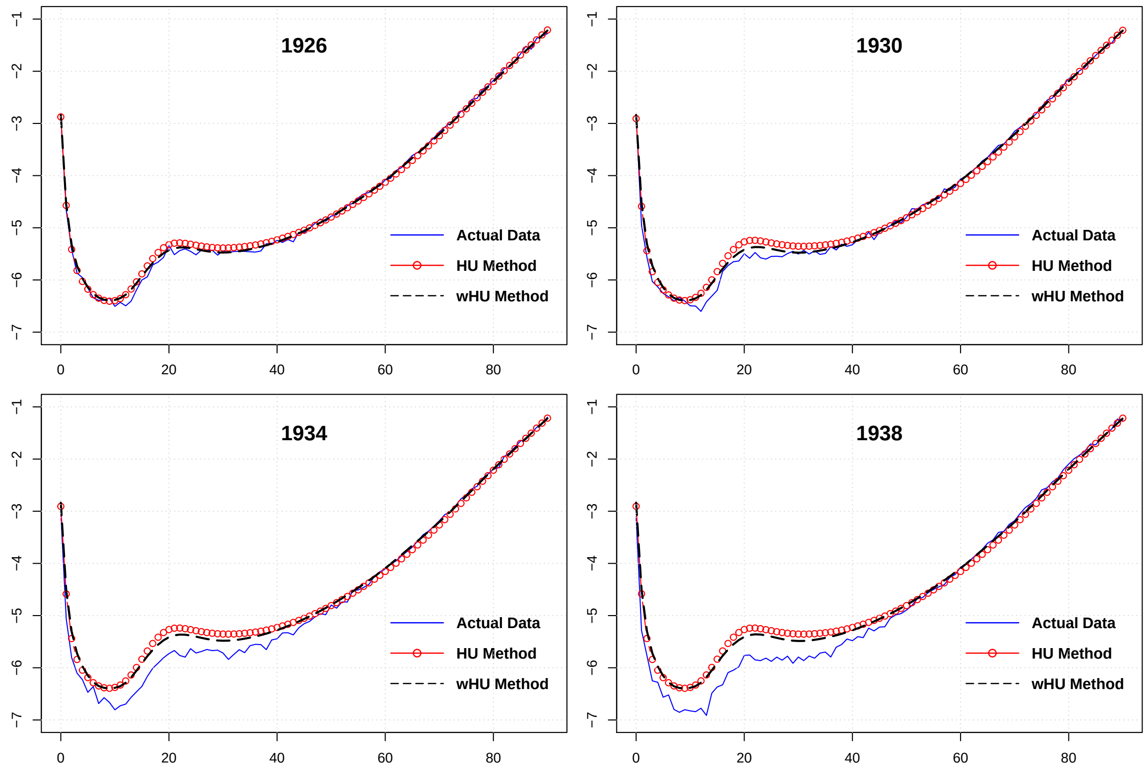

Figure 3.3: Mortality forecast of Sweden after the WWI and Spanish flu based on the training set $(1872-1925)$ for the forecasting horizon of 13 years.

Regarding Sweden, the wHU method offers slightly better estimations of the future log death rates throughout the whole testing horizon. Very high accuracy can be observed in the first years of the forecast by both methods, with slight overestimations of mortality for the ages between 15 and 30. In the year 1934, we can notice a larger gap between the actual and predicted data, while the age span increased to a range between early childhood and approximately 45. A similar range of ages is observable for the year 1938, but the overestimation rose even more.

Figure 3.4 below shows the curve of the French historical data in comparison to the constructed curves of mortality rates applying HU and wHU methods.

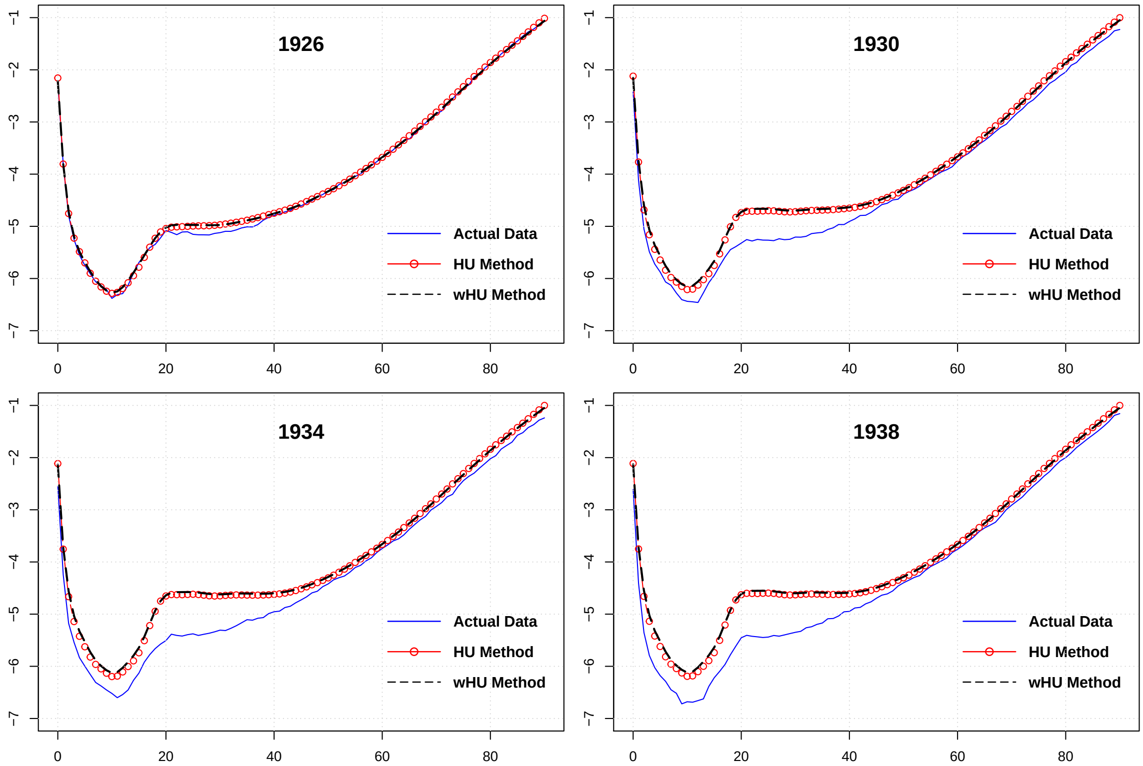

Figure 3.4: Mortality forecast of France after the WWI and Spanish flu based on the training set (1872 – 1925) for the forecasting horizon of 13 years.

Regarding France, the first forecasted year 1926 shows the tendency for overestimating the mortalities of the critical ages, which was confirmed in the following years. We can observe a gradual increase of the distance between the historical and forecasted log death rates across years, characteristic for the ages until approximately 45.

It can be concluded that HU and wHU methods do not make significantly different forecasts during the whole testing period. However, being influenced by the increased mortality during WWI and Spanish flu, both methods tend to overestimate death rates of the ages that recorded the largest mortality jumps in the outlying years. Also, the forecasts are less accurate as the time after the outlying years passes.

The package *demography* [9] offers variety of options. Except for mortality, we can compare the actual and forecasted life expectancies for each of the selected countries over a certain period. The comparison of the actual and forecasted life expectancies for Sweden and France using HU and wHU methods is presented in the Tables 3.2 and 3.3.

Table 3.2: Demographic comparison for Sweden and France after the WWI and Spanish flu using HU method based on the training set (1872 – 1925).

Sweden - Life expectancy (years)			
Year	Actual	Forecast	Error (HU)
1926	62.747	62.424	-0.323
1927	61.555	62.465	0.909
1928	62.211	62.493	0.281
1929	62.296	62.510	0.214
1930	63.178	62.519	-0.658
1931	62.717	62.525	-0.192
1932	63.946	62.527	-1.419
1933	64.813	62.527	2.286
1934	64.968	62.527	-2.441
1935	64.858	62.526	-2.332
1936	64.606	62.525	-2.080
1937	64.620	62.524	-2.096
1938	65.545	62.523	-3.022
Mean	63.697	62.509	-1.188

France - Life expectancy (years)			
Year	Actual	Forecast	Error (HU)
1926	53.949	52.792	-1.157
1927	55.740	52.133	-3.606
1928	55.405	51.642	-3.763
1929	54.226	51.278	-2.948
1930	56.814	51.008	-5.806
1931	56.909	50.809	-6.100
1932	57.231	50.663	-6.568
1933	57.664	50.556	-7.108
1934	58.328	50.477	-7.851
1935	58.301	50.420	-7.881
1936	58.778	50.378	-8.400
1937	59.144	50.348	-8.796
1938	58.959	50.326	-8.633
Mean	57.034	50.987	-6.048

The actual life expectancies for Sweden and France in the range between 1926 and 1938 indicate that the expected lifetime in Sweden was around nine years longer in the 1926. In 1938, that difference fell to approximately six and half years, mostly due to the large jump in the French expected lifetime. As shown in Fig. 3.3, the HU method slightly overestimates Swedish death rates and, therefore, underestimates the life expectancy for the average of 1.188 years.

As already mentioned, France had more outlying years during the period of the WWI and Spanish flu than Sweden. Accordingly, it is expected that the forecasts will provide larger measures of error, while the accuracy will be diminished. Based on Fig. 3.4, the HU method more notably overestimates French mortality rates and, therefore, underestimates the expected lifetime for the average of 6.048 years.

Table 3.3: Demographic comparison for Sweden and France after the WWI and Spanish flu using wHU method based on the training set (1872 – 1925).

Sweden - Life expectancy (years)				**France - Life expectancy (years)**			
Year	Actual	Forecast	Error (wHU)	Year	Actual	Forecast	Error (wHU)
1926	62.747	62.428	-0.319	1926	53.949	53.129	-0.820
1927	61.555	62.415	0.860	1927	55.740	52.324	-3.416
1928	62.211	62.409	0.198	1928	55.405	51.677	-3.727
1929	62.296	62.405	0.109	1929	54.226	51.222	-3.005
1930	63.178	62.402	-0.776	1930	56.814	50.868	-5.946
1931	62.717	62.398	-0.318	1931	56.909	50.594	-6.316
1932	63.946	62.395	-1.551	1932	57.231	50.405	-6.827
1933	64.813	62.392	-2.421	1933	57.664	50.261	-7.403
1934	64.968	62.388	-2.580	1934	58.328	50.151	-8.177
1935	64.858	62.385	-2.473	1935	58.301	50.075	-8.226
1936	64.606	62.382	-2.223	1936	58.778	50.019	-8.759
1937	64.620	62.379	-2.241	1937	59.144	49.975	-9.168
1938	65.545	62.376	-3.169	1938	58.959	49.946	-9.013
Mean	63.697	62.396	-1.300	**Mean**	57.034	50.819	-6.216

As is the case for the HU method, the considerably better accuracy is observable for the wHU method when the forecast is performed on the less outlying data set, such as Swedish. Regarding life expectancies, the wHU method records slightly larger mean errors for both the Swedish and French data sets. This phenomenon can be explained by the wHU method's weights, which give additional importance to the more recent historical data. Since those recent years were outlying, they additionally influenced the forecast. This problem can be very intriguing for further investigation, i.e. evaluating forecasting accuracy depending on the position of the outliers in the training set.

3.2.2 WWII

The second outlying period that is going to be tested is the period of the Second World War, which lasted from 1939 to 1945. The forecast of the mortality rates in the years after that critical period asks for one important decision. The question is whether to make the training set shorter by removing the outlying period of the WWI and Spanish flu from the training set. The other option is having a longer training set but including two world wars and Spanish flu inside. We want to investigate whether the forecast gets more or less accurate if one of the two outlying periods is not considered, but for the price of a significantly shorter training set. Therefore, both training sets will be tested.

Let the training set include only the years after the 1923, i.e. from 1924 until 1950. This year's span is remarkably small and we should be aware of its potential influence on the forecast. The forecasting horizon is a 40-year span that lasts from 1951 until 1990.

The HU method - For the selected Swedish data set and $K = 4$ principal components, the explained percentage variation due to basis functions is 96.49%, 1.53%, 0.71% and 0.42%, respectively. That gives the sum of 99.15%, while 0.85% of the variation is left unexplained.

For the selected French data set and $K = 6$ principal components, the basis functions explain 87.53%, 8.37%, 2.23%, 0.84%, 0.43% and 0.35%, respectively. That sums to the total of 99.75%, while only 0.25% of the variation is left unexplained.

The wHU method - Considering the selected Swedish data set for $K = 4$ principal components, the explained percentage variation due to basis functions is 97.72%, 0.99%, 0.46% and 0.27%, respectively. That means that the total is 99.44%, while only 0.56% of the variation is left unexplained.

Considering the selected French data set for $K = 6$ principal components, the basis functions explain 88.08%, 8.55%, 1.84%, 0.69%, 0.35% and 0.29%, respectively. That sums to the total of 99.80%, while only 0.20% of the variation is left unexplained. As was the case in the previous forecasts, for the equal number of principal components, the wHU method explains more percentage variation for both Sweden and France, compared to the HU method. In the Table 3.4 below, we present numerical values of each of the explained error measures for the Second World War forecast based on the shorter training set.

Table 3.4: Mean errors of the in-sample forecast of the WWII for Sweden and France based on the training set $(1924 - 1950)$.

	Sweden					**France**				
	ME	MAE	MSE	MPE	MAPE	ME	MAE	MSE	MPE	MAPE
HU	-0.0769	0.1380	0.0435	-0.0579	0.1219	0.6912	0.6921	0.7283	1.2800	1.2809
wHU	-0.0739	0.1433	0.0516	-0.0522	0.1246	0.6834	0.6846	0.7230	1.2688	1.2699

This table of errors provides interesting information. Regarding the Swedish data set, the measured errors of the forecasts made by the HU and wHU methods do not agree on which method is more accurate. The HU method minimized MAE, MSE and MAPE, while the wHU method minimized ME and MPE. Regarding the French data set, the wHU method performs with better accuracy since it minimizes all five error measures.

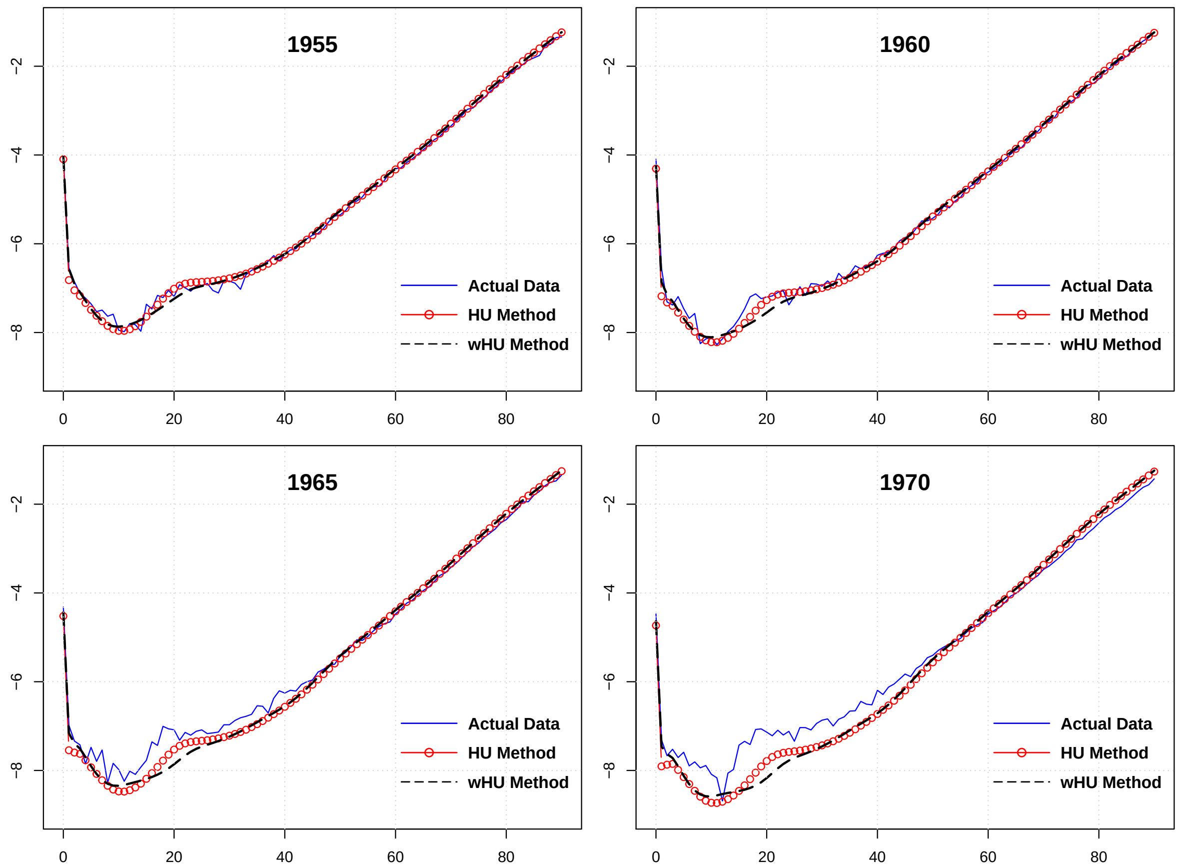

Figure 3.5: Mortality forecast of Sweden after the WWII for the training set (1924 – 1950) and forecasting horizon of 20 years.

For the Swedish data set and the year 1955, the forecasts of the HU and wHU methods nicely fit the historical mortality curve in general, with only small deviations for the ages of early childhood until approximately 30. The remaining ages perfectly fit the actual data. So far, the forecasts by both applied methods never showed a tendency for underestimation of the mortality. However, based on the shorter training set, they do. For the year 1960, the slight underestimation of mortality is present for the ages between approximately 2 and 7 for both methods and approximately 15 to 20 for the HU method and 15 to 23 for the wHU method. Besides that, the death rates for the remaining ages are precisely predicted. The two remaining graphs, years 1965 and 1970, indicate that forecast accuracy drops as the time horizon extends. Both methods tend to underestimate mortality rates for the ages until approximately 45.

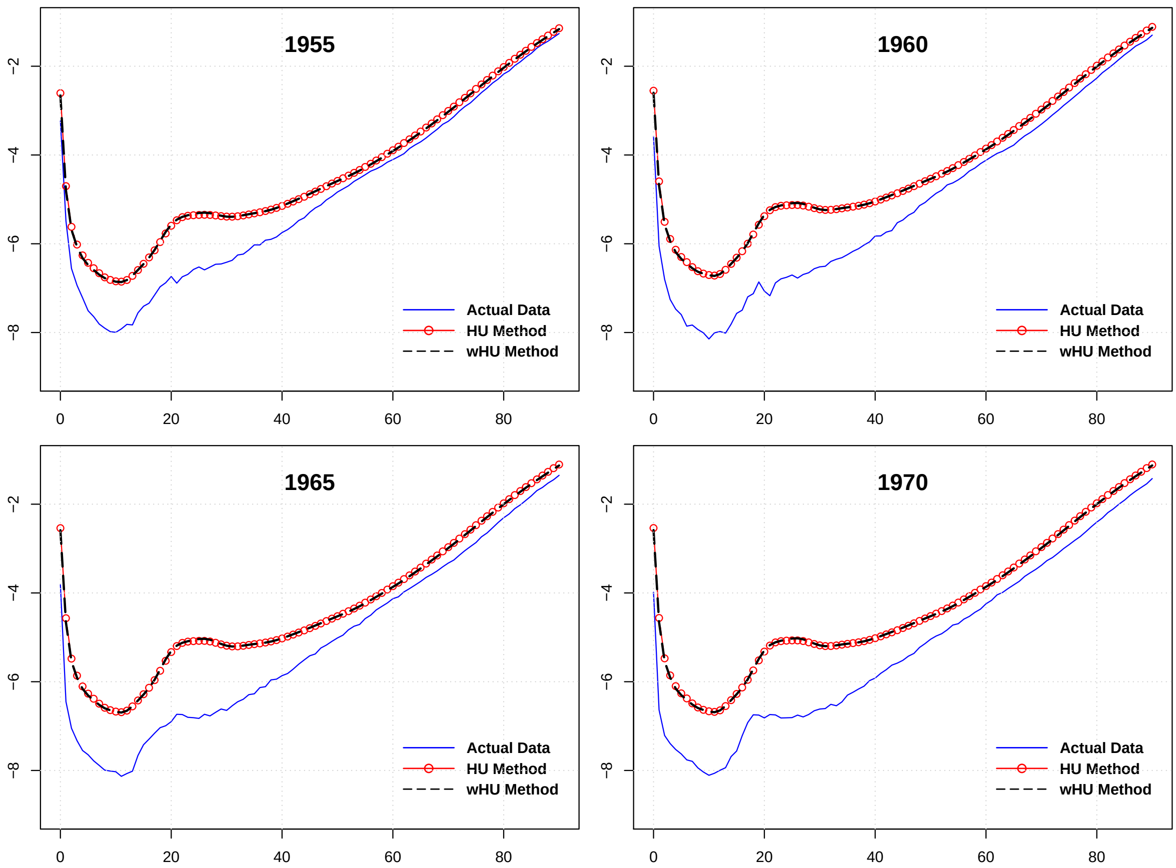

Figure 3.6: Mortality forecast of France after the WWII for the training set (1924 – 1950) and forecasting horizon of 20 years.

All of the mortality forecasts on the French data set are showing similar results. The log death rates are constantly overestimated, most probably due to the combination of the large mortality shocks that France experienced in the WWII and short training set.

Based on the Table 3.5, the life expectancy in Sweden in the period between 1951 and 1970 is forecasted with large accuracy. The HU method precisely estimates the lifetime, making no individual error to be larger than half a year, with the average error of 0.033 years.

Based on Fig. 3.6, the HU method is constantly overestimating mortality rates. Therefore, the expected lifetime is underestimated for the average of 10.012 years, which is a quite large deviation from historical data.

The results presented in the Table 3.5 bring us to the conclusion that reducing the length of the training set has a positive impact on the HU forecast of the life expectancy in the country that did not experience significant mortality shocks. On the other side, if the country experienced several mortality shocks such as France during WWII, the forecast gets extremely inaccurate.

Table 3.5: Demographic comparison for Sweden and France after the WWII using HU method for the training set $(1924-1950)$.

Sweden - Life expectancy (years)				France - Life expectancy (years)			
Year	Actual	Forecast	Error (HU)	Year	Actual	Forecast	Error (HU)
1951	71.378	71.340	0.022	1951	66.108	64.573	-1.535
1952	71.867	71.669	-0.198	1952	67.412	63.143	-4.269
1953	71.934	71.903	-0.031	1953	67.333	62.143	-5.190
1954	72.376	72.128	-0.248	1954	68.217	61.317	-6.899
1955	72.626	72.337	-0.288	1955	68.465	60.647	-7.818
1956	72.690	72.537	-0.154	1956	68.482	60.169	-8.314
1957	72.513	72.728	0.215	1957	68.920	59.799	-9.121
1958	73.179	72.912	-0.266	1958	70.140	59.510	-10.630
1959	73.407	73.091	-0.316	1959	70.212	59.302	-10.910
1960	73.057	73.264	0.207	1960	70.399	59.145	-11.254
1961	73.522	73.432	-0.090	1961	71.017	59.026	-11.991
1962	73.370	73.595	0.224	1962	70.503	58.939	-11.564
1963	73.583	73.753	0.170	1963	70.354	58.874	-11.479
1964	73.750	73.907	0.156	1964	71.337	58.826	-12.511
1965	73.911	74.056	0.145	1965	71.149	58.790	-12.360
1966	74.145	74.201	0.056	1966	71.573	58.763	-12.810
1967	74.201	74.342	0.141	1967	71.548	58.744	-12.805
1968	74.041	74.479	0.438	1968	71.537	58.729	-12.808
1969	74.153	74.613	0.459	1969	71.255	58.718	-12.536
1970	74.726	74.743	0.017	1970	72.156	58.710	-13.446
Mean	73.221	73.254	0.033	**Mean**	69.906	59.893	-10.012

Based on the Table 3.6, the wHU forecast offers even better results in terms of the average error that equals -0.147. Therefore, the life expectancy in Sweden in the period between 1951 and 1970 is again forecasted with large accuracy.

As was the case with the HU forecast, the wHU method is constantly overestimating death rates. Therefore, the expected lifetime is underestimated for the average of 9.685 years, which is still very far from the actual data.

Based on the results presented in Table 3.6, it can be concluded that shortening the training set improves the accuracy of the life expectancy forecast using the wHU method for the data set without the large mortality shocks. Oppositely, if the data set includes significant jumps in mortality, the forecast becomes strongly inaccurate.

Table 3.6: Demographic comparison for Sweden and France after the WWII using wHU method for the training set $(1924 - 1950)$.

Sweden - Life expectancy (years)			
Year	Actual	Forecast	Error (wHU)
1951	71.378	71.366	-0.012
1952	71.867	71.580	-0.287
1953	71.934	71.787	-0.147
1954	72.376	71.989	-0.387
1955	72.626	72.184	-0.442
1956	72.690	72.373	-0.318
1957	72.513	72.556	0.042
1958	73.179	72.734	-0.445
1959	73.407	72.906	-0.501
1960	73.057	73.074	0.017
1961	73.522	73.236	-0.286
1962	73.370	73.394	0.024
1963	73.583	73.547	-0.036
1964	73.750	73.696	-0.054
1965	73.911	73.841	-0.071
1966	74.145	73.981	-0.164
1967	74.201	74.118	-0.083
1968	74.041	74.251	0.209
1969	74.153	74.380	0.227
1970	74.726	74.506	-0.220
Mean	73.221	73.075	-0.147

France - Life expectancy (years)			
Year	Actual	Forecast	Error (wHU)
1951	66.108	64.827	-1.281
1952	67.412	63.512	-3.901
1953	67.333	62.457	-4.876
1954	68.217	61.628	-6.589
1955	68.465	60.986	-7.479
1956	68.482	60.495	-7.987
1957	68.920	60.122	-8.797
1958	70.140	59.842	-10.298
1959	70.212	59.632	-10.580
1960	70.399	59.475	-10.925
1961	71.017	59.357	-11.659
1962	70.503	59.270	-11.233
1963	70.354	59.206	-11.148
1964	71.337	59.158	-12.179
1965	71.149	59.122	-12.027
1966	71.573	59.096	-12.477
1967	71.548	59.076	-12.472
1968	71.537	59.062	-12.476
1969	71.255	59.051	-12.204
1970	72.156	59.043	-13.113
Mean	69.906	60.221	-9.685

3.2.3 WWII with the Longer Training Set

Let this longer training set cover the years from 1872 until 1950, including them. Afterwards, let the testing set cover the interval of the succeeding 40 years - from 1951 until 1990, including them.

The HU method - For the selected Swedish data set, the optimal number of principal components is $K = 4$. Accordingly, the explained percentage variation due to basis functions is 93.95%, 4.25%, 1.09% and 0.28%, respectively. That gives the sum of 99.57%, while only 0.43% of the variation is left unexplained.

For the selected French data set, the optimal number of principal components is $K = 6$. Therefore, the basis functions explain 81.57%, 14.97%, 1.34%, 0.61%, 0.44% and 0.36%, respectively. That sums to the total of 99.29%, while only 0.71% of the variation is left unexplained.

The wHU method - Considering the selected Swedish data set for $K = 4$ principal components, the explained percentage variation due to basis functions is 98.08%, 1.38%, 0.33% and 0.09%, respectively. Therefore, the total is 99.88%, while only 0.12% is left unexplained.

Considering the selected French data set for $K = 6$ principal components, the basis functions explain 92.09%, 6.50%, 0.56%, 0.25%, 0.18% and 0.14%, respectively. That sums to the total of 99.72%, while only 0.28% of the variation is left unexplained. Once again, for the equal number of principal components, the wHU method explains more percentage variation for both countries in comparison to the HU method. In the Table 3.7, numerical values of each of the explained error measures for the Second World War forecast are presented.

Table 3.7: Mean errors of the in-sample forecast of the WWII for Sweden and France based on the training set $(1872 - 1950)$.

	Sweden					**France**				
	ME	MAE	MSE	MPE	MAPE	ME	MAE	MSE	MPE	MAPE
HU	0.0784	0.0986	0.0134	0.0855	0.1046	0.2495	0.2540	0.0978	0.3081	0.3124
wHU	0.0435	0.0719	0.0094	0.0484	0.0751	0.2340	0.2397	0.0917	0.2891	0.2947

Based on the calculations provided in the table above, it can be concluded that the wHU method decreases the error in 10 out of 10 cases, making no di□erence in the country or the type of the error. This indicates more precise forecasts of mortalities in the application of the wHu method, compared to the standard HU method.

Figure 3.7 shows the comparison of the mortality forecasts after WWII for Sweden, based on the shorter and longer training sets. Figure 3.8 presents the curve of the French historical data in comparison to the constructed curves of mortality rates applying HU and wHU methods. On all graphs, the x-axis represents age and the y-axis represents log mortality rate.

Regarding Sweden, it can be observed that both forecasting methods very accurately fit the shape of the historical data. Swedish mainly neutral position during the WWII secured the absence of any significant variation in the log death rates. For the years 1955 and 1960, HU and wHU methods slightly overestimate log mortalities for ages between approximately 20 and 32, probably due to the fact that this age group su□ered the most during war times. Until 1965, both methods are successfully smoothing the forecasts even more. Starting from 1970, a very slight overestimation of the mortality for ages older than 65 to 70 can be observed.

After analyzing the plots, it can be concluded that there is no significant di□erence between the HU and wHU methods in the case when the data set does not contain large mortality shocks. In all presented graphs, the slight deviations are happening for ages younger than approximately 30. However, both HU and wHU methods almost perfectly forecast the mortality rates for the remaining ages. By comparing the obtained forecasts of mortality after the Second World War in Sweden based on the training sets $(1972 - 1950)$ and $(1924 - 1950)$, we observe that shortening the training set deteriorated forecasts to a certain extent. It is especially visible after the first 20 years of the forecast. Therefore, it can be concluded that for the country without many large outliers, the mortality rates are more accurately forecasted if the training set is longer, even for the price of the inclusion of another outlying period.

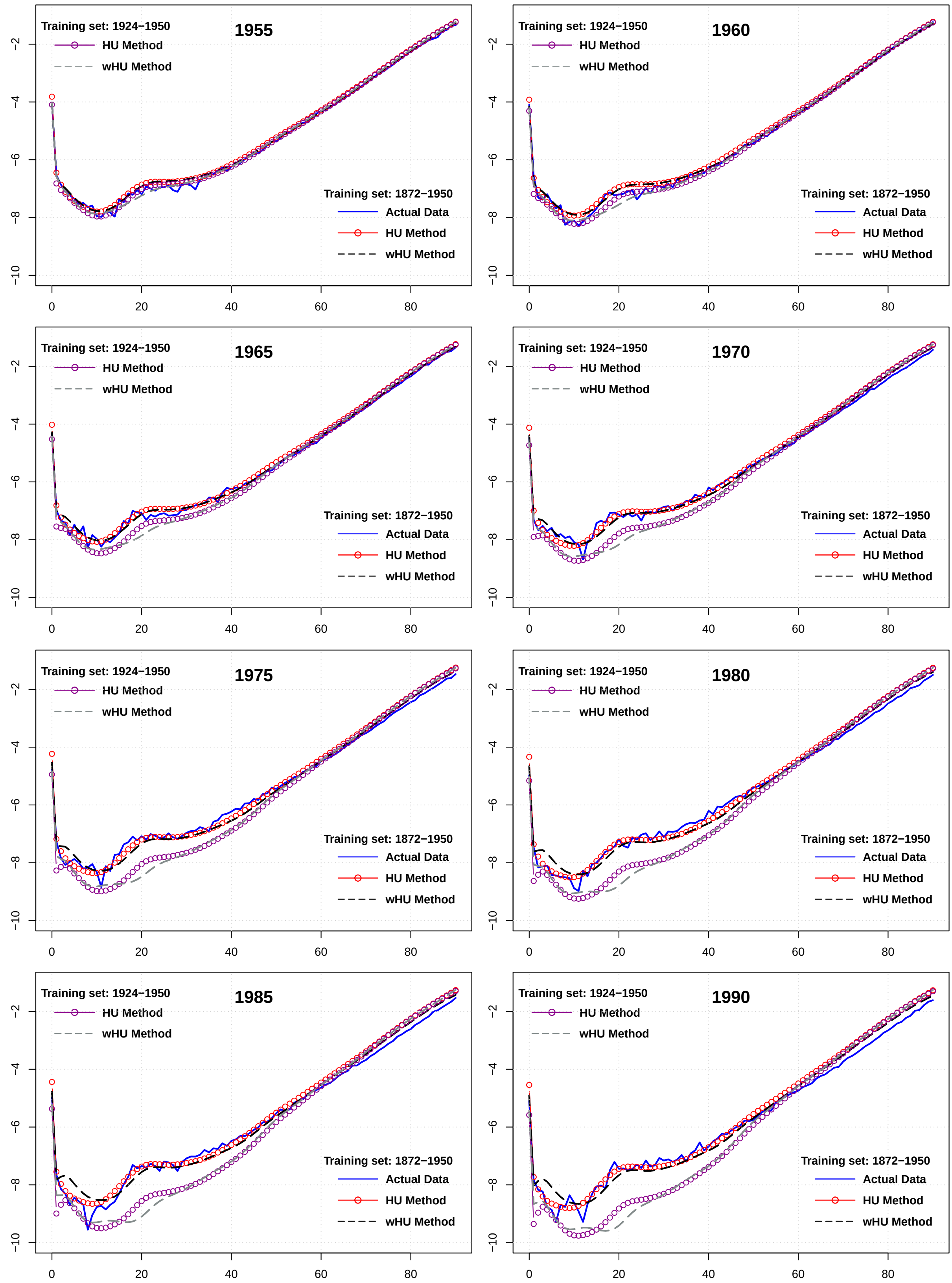

Figure 3.7: Comparison of the mortality forecasts of Sweden after the WWII for the shorter (1924 – 1950) and longer (1872 – 1950) training sets and forecasting horizon of 40 years.

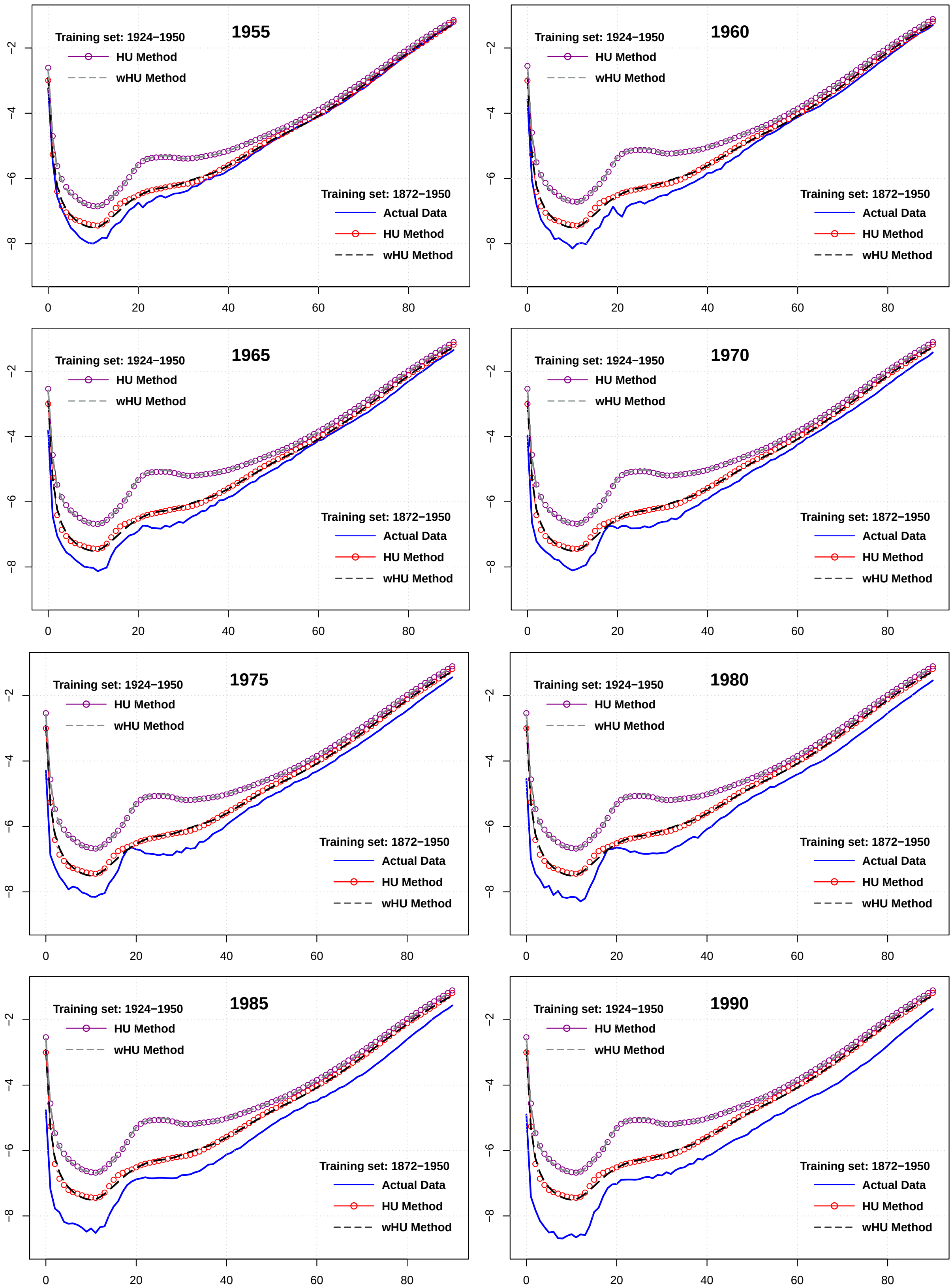

Figure 3.8: Comparison of the mortality forecasts of France after the WWII for the shorter (1924 – 1950) and longer (1872 – 1950) training sets and forecasting horizon of 40 years.

Regarding France, it can be observed that the HU and wHU methods have almost identical predictions of mortality rates. For the year 1955, starting from a very young age, the forecasting methods overestimate mortalities to approximately 35, while the mortality rates of all remaining ages almost perfectly fit the historical data. Depending on the time distance from the outlying years, both HU and wHU methods tend to overestimate log death rates for a wider range of ages. Precisely, for 1960, the log death rates of all ages older than approximately 50 almost perfectly fit the actual data. The year 1965, in addition, presents a slight deviation from the historical data even for the ages between 65 and 85, while the year 1970 shows a constant overestimation of the mortality across all ages. This trend of death rate overestimations continues and increases with each following year.

No significant difference between the two methods has been noticed in the plots. However, in the case when the data set contains some mortality shocks as French does, it is observable that HU and wHU methods tend to generate larger forecasting errors as the time from the outlying period passes.

However, for the historical data filled with outliers, the graphs offer a quality insight into how superior the forecasts are when a longer training set is applied. The forecasts based on the training set (1924 – 1950) are very inaccurate compared to the forecasts based on the training set (1872 – 1950). The overestimation in mortalities for ages until approximately 45 is significantly abbreviated. Therefore, the obtained results indicate that the HU and wHU methods can relatively successfully handle the data sample with large outliers such as French as the forecasting horizon is shorter. But for all tested years, the forecasts are more accurate for the extended training set, even for the price of inclusion of the additional outlying periods.

Regarding life expectancy, the Table 3.8 below presents much better accuracy when making the HU forecast based on the data set with some or no large outliers (Sweden) than based on the data set with large outliers (France). By comparing the obtained forecasts of mortality after the Second World War in France based on the training sets (1972 – 1950) and (1924 – 1950), we notice that shortening the training set negatively influenced the accuracy of the forecasts.

The historical life expectancies for Sweden and France in the period 1951 – 1970 suggest that the expected lifetime in Sweden was about five years longer in the first years of that period. In 1970, that difference fell to approximately two and half years, which is a significant improvement for France. As shown in Fig. 3.7, the HU method tends to slightly overestimate Swedish mortality rates, underestimating the expected lifetime for the average of 1.22 years.

Due to its active participation in the Second World War, France experienced greater mortality shocks than Sweden. Therefore, it could be expected that the forecast of mortality, as well as the life expectancy, be accompanied by a larger forecasting error. Based on the Fig. 3.8, the HU method constantly overestimates French mortality rates and, therefore, underestimates the expected lifetime for the average of 3.525 years.

As was the case for the HU method, much better accuracy is present when making the wHU forecast on the data set with some or no large outliers (Sweden) than on the data set with large outliers (France). This can be seen in the Table 3.9.

Table 3.8: Demographic comparison for Sweden and France after the WWII using HU method for the training set (1872 – 1950) and forecasting horizon of 20 years.

Sweden - Life expectancy (years)			
Year	Actual	Forecast	Error (HU)
1951	71.378	70.964	-0.414
1952	71.867	71.059	-0.808
1953	71.934	71.166	-0.768
1954	72.376	71.279	-1.097
1955	72.626	71.394	-1.232
1956	72.690	71.509	-1.181
1957	72.513	71.625	-0.889
1958	73.179	71.739	-1.440
1959	73.407	71.852	-1.555
1960	73.057	71.964	-1.093
1961	73.522	72.074	-1.448
1962	73.370	72.183	-1.187
1963	73.583	72.291	-1.292
1964	73.750	72.397	-1.353
1965	73.911	72.502	-1.409
1966	74.145	72.605	-1.540
1967	74.201	72.707	-1.494
1968	74.041	72.807	-1.234
1969	74.153	72.907	-1.246
1970	74.726	73.005	-1.721
Mean	73.221	72.001	-1.220

France - Life expectancy (years)			
Year	Actual	Forecast	Error (HU)
1951	66.108	66.492	0.384
1952	67.412	66.458	-0.954
1953	67.333	66.465	-0.868
1954	68.217	66.412	-1.805
1955	68.465	66.404	-2.061
1956	68.482	66.395	-2.087
1957	68.920	66.386	-2.534
1958	70.140	66.378	-3.762
1959	70.212	66.371	-3.841
1960	70.399	66.365	-4.034
1961	71.017	66.360	-4.657
1962	70.503	66.356	-4.147
1963	70.354	66.353	-4.001
1964	71.337	66.350	-4.987
1965	71.149	66.347	-4.802
1966	71.573	66.346	-5.227
1967	71.548	66.344	-5.204
1968	71.537	66.343	-5.194
1969	71.255	66.342	-4.913
1970	72.156	66.341	-5.815
Mean	69.906	66.380	-3.525

As it can be observed from the Fig. 3.7, the wHU method also overestimates Swedish mortality rates but the curve is located closer to the curve of the actual data. This indicates higher accuracy which is confirmed by underestimating the expected lifetime for the average of 0.409 years.

On the other side, the wHU method does not provide improved forecasted life expectancy for the French data set. In fact, the mean error increased but only at the second decimal which can not be considered significant. Observing the Fig. 3.8, the wHU method steadily overestimates French mortality rates and underestimates the life expectancy for an average of 3.582 years. This can be observed in the Table 3.9 below.

Table 3.9: Demographic comparison for Sweden and France after the WWII using wHU method for the training set (1872 – 1950) and forecasting horizon of 20 years.

Sweden - Life expectancy (years)			
Year	Actual	Forecast	Error (wHU)
1951	71.378	71.405	0.027
1952	71.867	71.627	-0.240
1953	71.934	71.812	-0.122
1954	72.376	71.971	-0.405
1955	72.626	72.118	-0.508
1956	72.690	72.258	-0.432
1957	72.513	72.394	-0.119
1958	73.179	72.528	-0.651
1959	73.407	72.660	-0.747
1960	73.057	72.790	-0.267
1961	73.522	72.918	-0.604
1962	73.370	73.045	-0.325
1963	73.583	73.170	-0.413
1964	73.750	73.293	-0.457
1965	73.911	73.415	-0.496
1966	74.145	73.536	-0.609
1967	74.201	73.655	-0.546
1968	74.041	73.773	-0.269
1969	74.153	73.889	-0.264
1970	74.726	74.004	-0.722
Mean	73.221	72.813	-0.409

France - Life expectancy (years)			
Year	Actual	Forecast	Error (wHU)
1951	66.108	66.305	0.197
1952	67.412	66.314	-1.098
1953	67.333	66.319	-1.014
1954	68.217	66.322	-1.894
1955	68.465	66.324	-2.141
1956	68.482	66.325	-2.157
1957	68.920	66.325	-2.595
1958	70.140	66.326	-3.814
1959	70.212	66.326	-3.886
1960	70.399	66.326	-4.074
1961	71.017	66.326	-4.691
1962	70.503	66.326	-4.177
1963	70.354	66.326	-4.028
1964	71.337	66.326	-5.011
1965	71.149	66.326	-4.823
1966	71.573	66.326	-5.247
1967	71.548	66.326	-5.242
1968	71.537	66.326	-5.211
1969	71.255	66.326	-4.929
1970	72.156	66.326	-5.830
Mean	69.906	66.324	-3.582

3.3 Out-of-Sample Forecasting

The in-sample forecasting provided some notable information that has to be considered in the process of the out-of-sample forecasting. The forecast accuracy highly depends on the number and intensity of the outlying years. The forecasted mortalities tend to be overestimated as the time after the outlying period passes.

The countries whose mortality rates are going to be forecasted can be selected according to many different criteria. The first and most important one is availability of the most recent historical data. To remain in harmony with the in-sample forecast presented in the section 3.2, the Human Mortality Database continues to be the source of the actual mortality rates. The intention was to continue performing forecasts for Sweden and France, but French mortality in HMD has been updated only until 2019, which does not provide enough information about period of the expansion of Covid-19 pandemics. The database contains Swedish mortality for 2020, so Sweden is the first country selected to take the part of the out-of-sample forecasting.

The only country that has available mortality for the year 2021 in the moment of writing is Denmark. Therefore, providing the most of the useful recent historical data, Denmark is the second selected member of the out-of-sample forecasting. The country whose population suffered strong hits by pandemics of Covid-19 is Spain. Representing one large outlier, Spain is the third chosen member. Finally, Japan is a good representative of the country outside of Europe whose mortality rates are going to be forecasted. According to the World Health Organization (WHO) [18], Japan is the country with the longest life expectancy in the world, which is additional reason why this forecast might be interesting to preform.

The training set for all countries covers the length of the 70 years. For Denmark, it lasts from 1952 until 2021, while for all other countries it starts at 1951 and ends at 2020. Regarding the principal components' optimal number selection, we continue the same reasoning applied in the in-sample forecast. Additionally, the graphs of the basis functions and associated coefficients are shown in Appendix B, while ARIMA forecasting coefficients obtained from the software and selection criteria are presented in Appendix C.

Table 3.10 represents the each country's summary of the explained variation for the corresponding optimal number of principal components.

Table 3.10: Explained percentage variation for Sweden, Denmark, Spain and Japan.

	Principal components					**Explained variation**
	1st	2nd	3rd	4th	5th	
Sweden	97.97%	0.99%	0.28%	0.21%	0.17%	99.62%
Denmark	96.86%	1.80%	0.44%	0.21%	0.17%	99.48%
Spain	98.11%	1.24%	0.30%	0.17%		99.82%
Japan	98.80%	0.88%	0.15%			99.83%

The following four graphs provide the mortality forecasts for each of the mentioned countries using the wHU method. In addition, the historical mortality curves of the previous 50 years are shown to emphasize the future trends. The fact that all graphs present the forecasts without drops in mortality is a good indicator that the wHU method works according to expectations. The interesting observation is that, except for the Swedish dataset, forecasts of the remaining countries show significant smoothing of the accident hump in the teenage years.

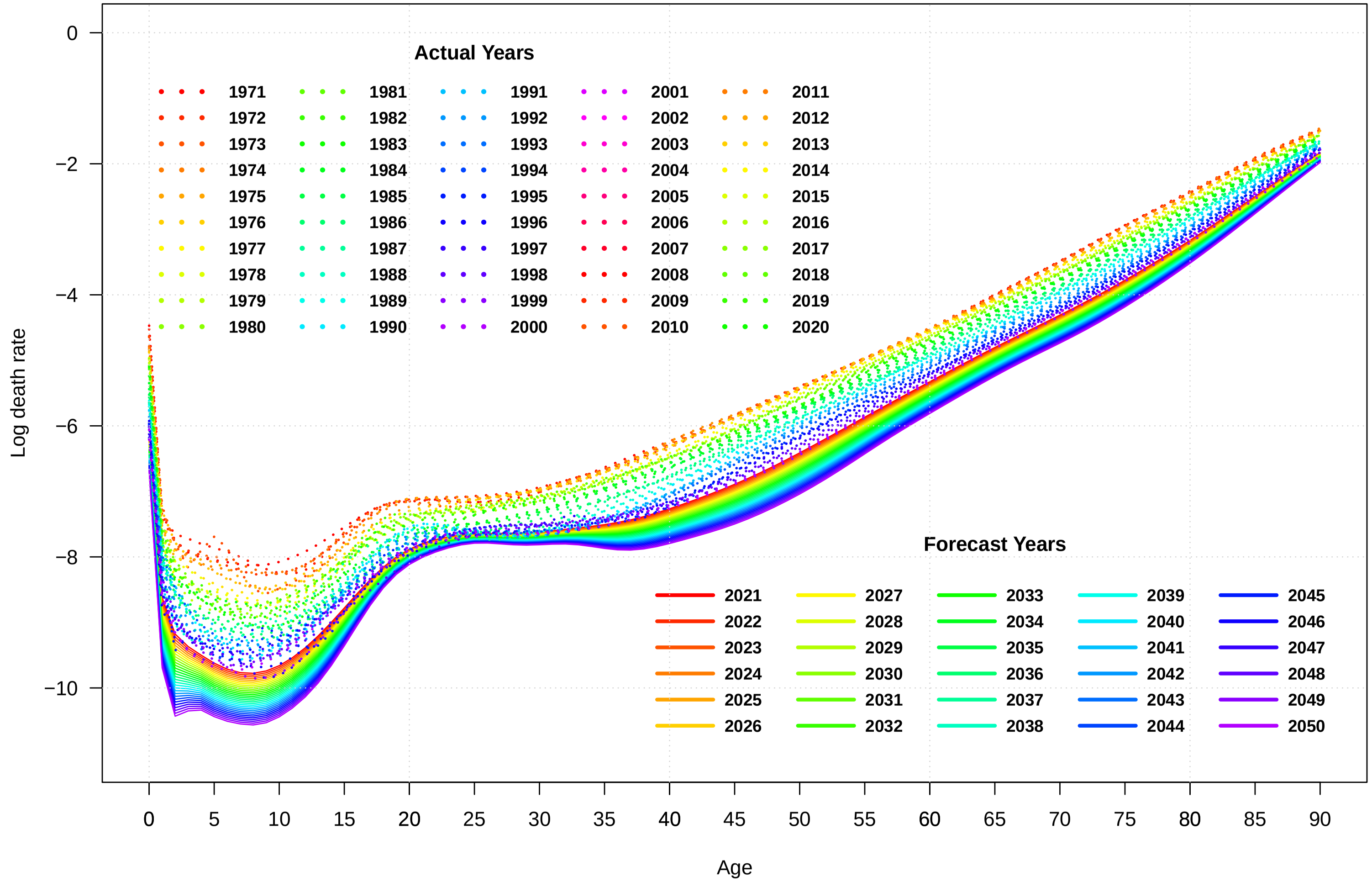

Figure 3.9: Mortality of Sweden: Actual (1971 – 2020) and Forecast (2021 – 2050).

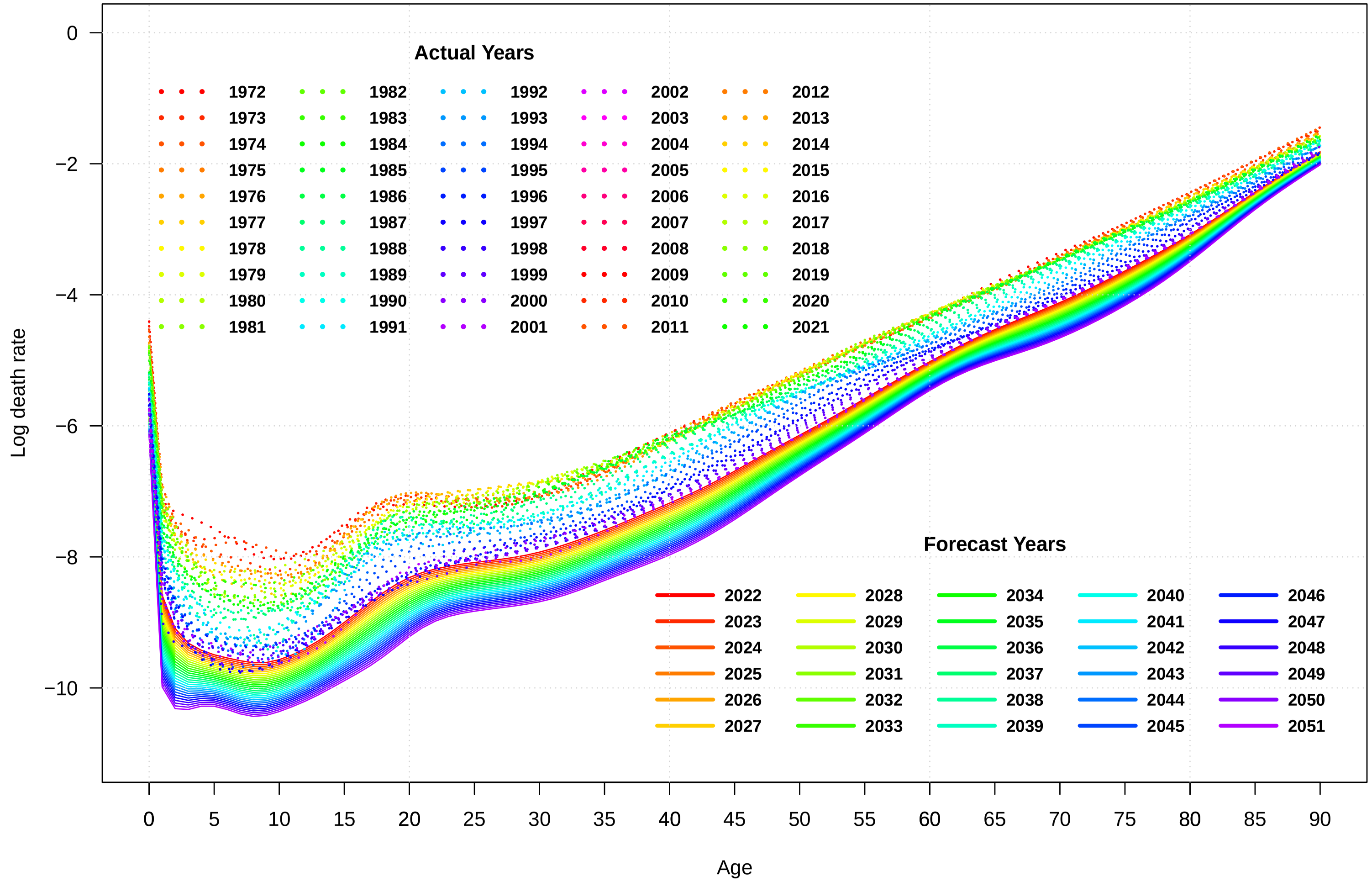

Figure 3.10: Mortality of Denmark: Actual (1972 – 2021) and Forecast (2022 – 2051).

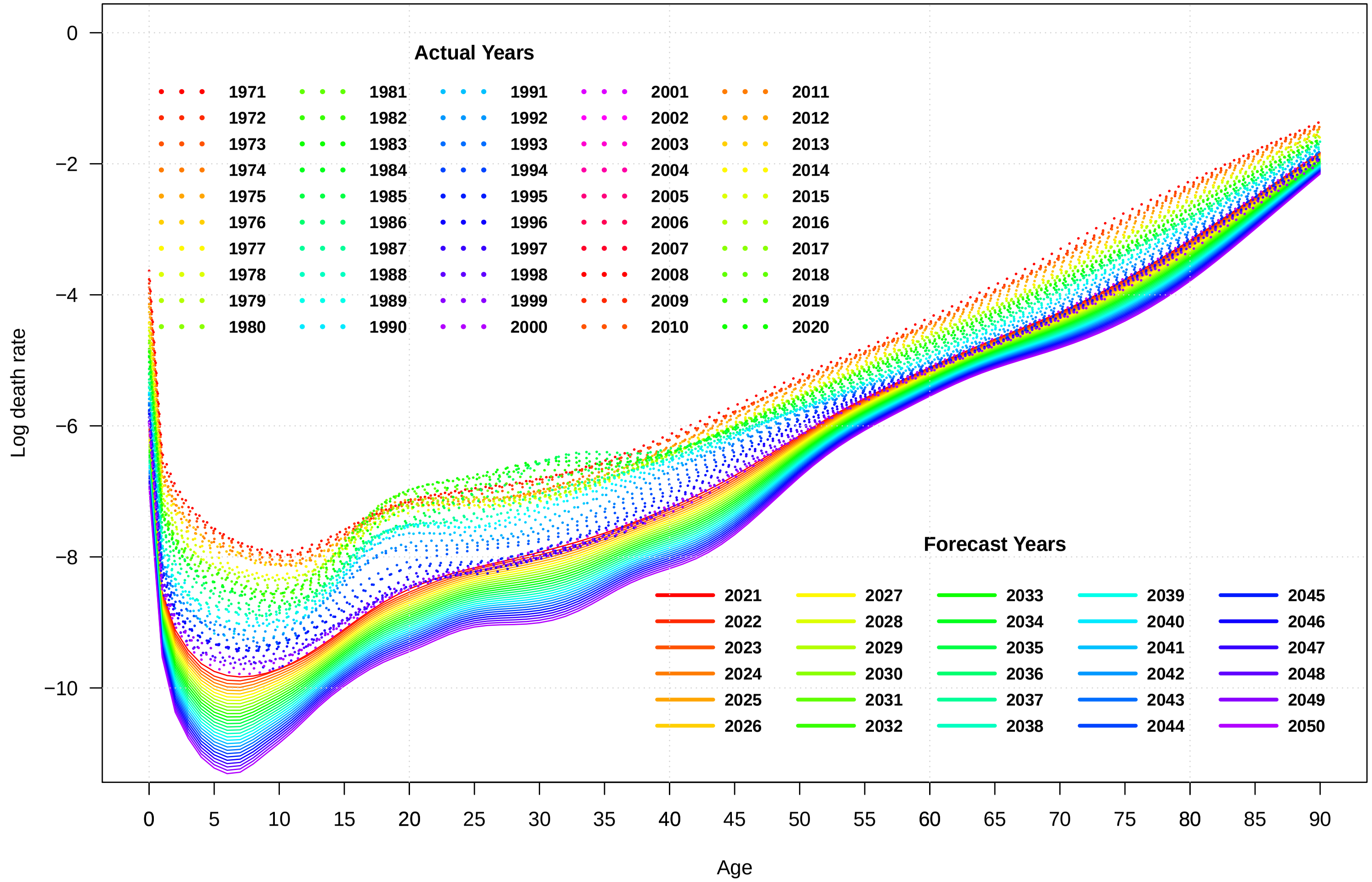

Figure 3.11: Mortality of Spain: Actual (1971 – 2020) and Forecast (2021 – 2050).

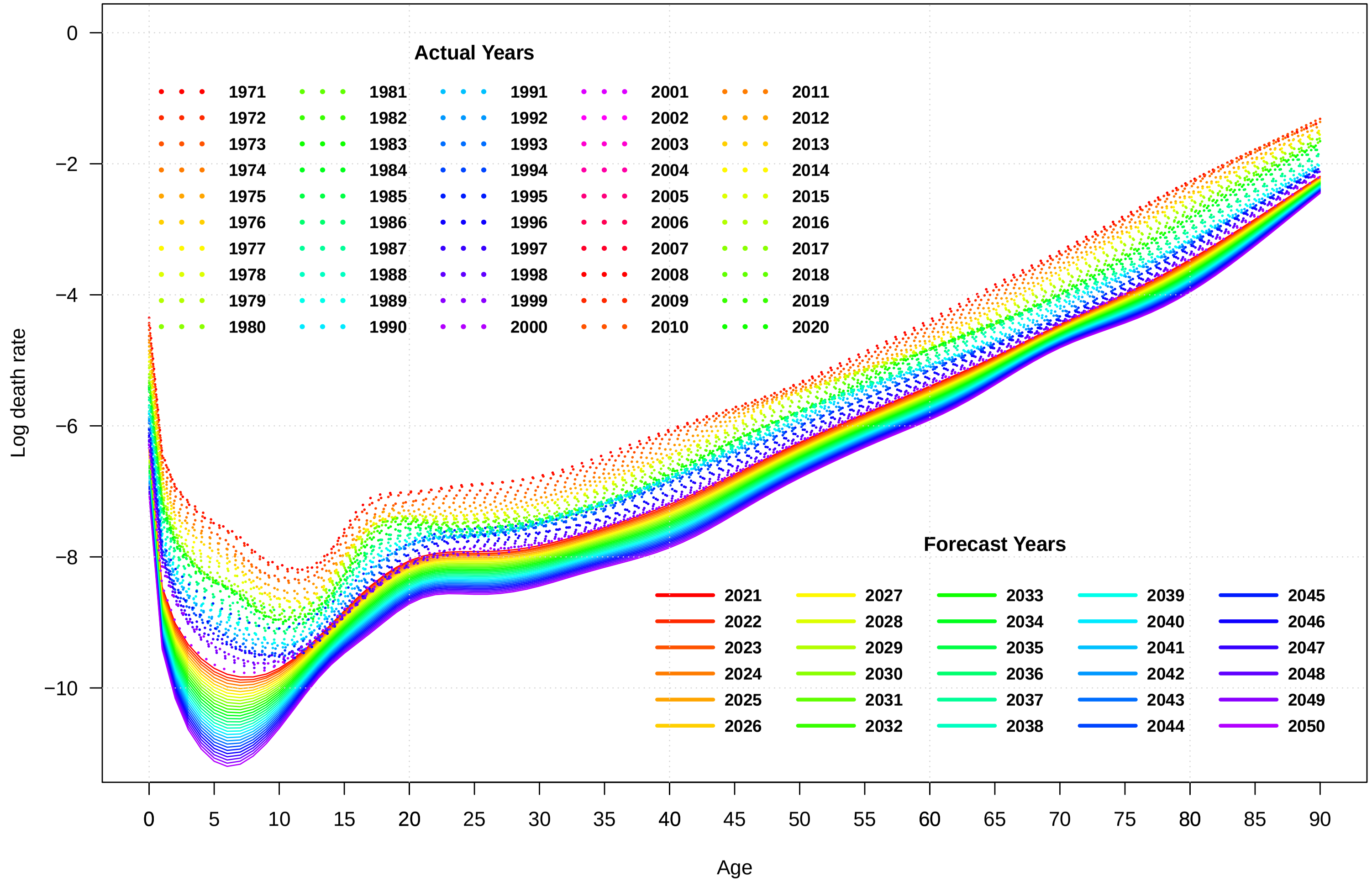

Figure 3.12: Mortality of Japan: Actual (1971 – 2020) and Forecast (2021 – 2050).

Chapter 4

Conclusion and Discussion

4.1 Conclusion

The forecasts presented in the previous Chapter provide numerous valuable conclusions regarding the efficiency of the HU and wHU methods when working with outliers. The R-package *demography* played an immense role in the simple but powerful analysis of the mortality data provided by HMD.

Throughout the whole testing period, the tested methods provide quite similar forecasts. Their common characteristic is overestimating mortalities associated with ages that recorded the largest mortality jumps during the outlying period. Nevertheless, the wHU method proved to be slightly, but consistently superior when working with the moderately outlying period such as in the Swedish case. Besides slight mortality overestimations for the range of ages between 15 and 30, the two methods recorded notable accuracy. However, the French data set filled with very turbulent periods regarding mortality, caused problems for both applied methods. Compared to the shorter French set (1924 – 1950) whose forecast of the mortality after WWII performed poorly, extending the training set to the range (1972 – 1950) recorded a significant improvement in accuracy. That brings the conclusion that the excessive shortening of the training set is much worse than the inclusion of another outlying period which provides a longer training set. Even if the historical data set consists of many outlying years, HU and wHU methods can relatively successfully handle them under the condition of a sufficiently long training set. Finally, in all cases, for both countries, the forecasts tend to be less accurate as the time distance from the outlying years rises.

Keeping in mind all conclusions obtained from the in-sample forecast, the post-Covid 19 mortality forecasts for four countries were performed. Regarding the out-of-sample forecast, no conclusions can be stated at the moment since only time can show how accurate the mortality forecasts are. Future research on this topic will undoubtedly provide more important answers and I hope that this thesis will contribute to this ever-lasting process.

4.2 Further Research

While working on this thesis, several ideas appeared to be very interesting as topics for future related work. As an author, my first desire was to forecast the mortality of my home country, Bosnia and Herzegovina. Bosnian history is full of conflicts accompanied by territorial integrations and disintegrations, thus being a potentially useful choice for forecasting mortality rates for countries that experienced mortality outliers. Unfortunately, such restless history is one of the reasons why it can be difficult to find a reliable source of data. The same problem is probably present in the majority of Balkan countries. If a trustworthy database exists and if it has enough data to contribute to the forecasts, exploring the various future demographic trends might be of great use to local governments.

Another country in the Balkan region with a rich history in the 20th century is Bulgaria. Namely, Bulgaria's participation in the First and Second Balkan Wars, WWI and WWII, may serve as an insightful example of the country that provides numerous outlying years of mortality. The main reason why mortality in this country was not analyzed and forecasted in this project is the fact that, at the moment of writing, the Human Mortality Database contains only data between WWII and before the outbreak of Covid-19.

It is important to distinguish the characteristics of localized and global wars. Definitely, they both are cause for more or less extreme increases in mortality rates in all participants of the conflict. However, the global wars, such as WWI and WWII affect the majority of countries in some way. Therefore, emigration can not be as helpful as it is in the more localized wars. For example, at the moment of writing, the war in Ukraine is happening and hundreds of thousands and even millions of Ukrainians are moving to Poland and other surrounding countries. Focusing mainly on Ukraine and Poland, the demographic pictures of both countries are extremely altered. Forecasting mortality rates in such conditions will be more challenging than usual.

Based on the analysis of shorter and longer training sets done in Section 3.2.3, one should extend the out-of-sample forecast done in Section 3.3 and include a longer training set. For example, one can consider the period of the Second World War, thus making the training set from 1935 until 2020.

Another interesting idea for future work would be to investigate how the position of the outlying years affects the HU and wHU forecasts. For example, in this thesis, the forecasts were performed based on the training sets in which the outlying years were near the end. But what if they were positioned in the middle or more towards the beginning of the observed period? Also, could the forecasts be more accurate if we modify the existing wHU model to be flexible in weights positioning? For instance, weighting the outlying period near the end of the training set might unnecessarily influence the forecasts after that critical period. On the other side, shifting the weights to the pre-outlying period might significantly improve the forecasts for the post-outlying period.

Bibliography

[1] David R Bellhouse. London plague statistics in 1665. *Journal of O cial Statistics*, 14(2):207, 1998.

[2] Heather Booth, Rob J Hyndman, Leonie Tickle, and Piet De Jong. Lee-carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic research*, 15:289–310, 2006.

[3] Heather Booth, John Maindonald, and Len Smith. Applying Lee-Carter under conditions of variable mortality decline. *Population studies*, 56(3):325–336, 2002.

[4] Piet De Jong and Leonie Tickle. Extending Lee–Carter mortality forecasting. *Mathematical Population Studies*, 13(1):1–18, 2006.

[5] David CM Dickson, Mary R Hardy, and Howard R Waters. *Actuarial mathematics for life contingent risks*. Cambridge University Press, 2013.

[6] Benjamin Gompertz. 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, 1825.

[7] Steven Haberman and Trevor A Sibbett. *History of Actuarial Science: Life insurance*, volume 5. Pickering & Chatto, 1995.

[8] Mia Hubert, Peter J Rousseeuw, and Sabine Verboven. A fast method for robust principal components with applications to chemometrics. *Chemometrics and intelligent laboratory systems*, 60(1-2):101–111, 2002.

[9] R Hyndman, H Booth, L Tickle, and J Maindonald. Package “demography”: Forecasting mortality, fertility, migration and population data. *R Package Version*, 1, 2019.

[10] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

[11] Rob J Hyndman and Heather Booth. Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 24(3):323–342, 2008.

[12] Rob J Hyndman, Anne B Koehler, Ralph D Snyder, and Simone Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of forecasting*, 18(3):439–454, 2002.

[13] Rob J Hyndman and Han Lin Shang. Forecasting functional time series. *Journal of the Korean Statistical Society*, 38(3):199–211, 2009.

[14] Rob J Hyndman and Md Shahid Ullah. Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956, 2007.

[15] Torsten Kleinow and Stephen J Richards. Parameter risk in time-series mortality forecasts. *Scandinavian Actuarial Journal*, 2017(9):804–828, 2017.

[16] Ronald Lee and Timothy Miller. Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38(4):537–549, 2001.

[17] Ronald D Lee and Lawrence R Carter. Modeling and forecasting US mortality. *Journal of the American statistical association*, 87(419):659–671, 1992.

[18] World Health Organisation. Life expectancy at birth (years), 2022. Retrieved from: `https://www.who.int/data/gho/data/indicators/indicator-details/GHO/life-expectancy-at-birth-(years)`. [Accessed: 28-April-2022].

[19] Anthony Kojo Ramos. Forecasting mortality rates using the weighted Hyndman-Ullah method, 2021.

[20] Arthur E Renshaw and Steven Haberman. A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and economics*, 38(3):556–570, 2006.

[21] Syazreen Shair, Sachi Purcal, and Nick Parr. Evaluating extensions to coherent mortality forecasting models. *Risks*, 5(1):16, 2017.

[22] Han Lin Shang, Heather Booth, and Rob J Hyndman. *A comparison of ten principal component methods for forecasting mortality rates*. Monash Univ., Department of Econometrics and Business Statistics, 2010.

[23] Han Lin Shang, Heather Booth, and Rob J Hyndman. Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research*, 25:173–214, 2011.

[24] Lenny Stoeldraijer, Coen van Duin, Leo van Wissen, and Fanny Janssen. Impact of di□erent mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the netherlands. *Demographic Research*, 29:323–354, 2013.

[25] Robert Tombs. *The Paris Commune 1871*. Routledge, 2014.

[26] Ruey S Tsay. *Analysis of financial time series*. John wiley & sons, 2005.

[27] Berkeley (USA) University of California and Max Planck Institute for Demographic Research (Germany). Human mortality database. Available at. `http://www.mortality.org`. [Accessed: 25-March-2022].

[28] Peter Wallström. *Evaluation of forecasting techniques and forecast errors: with focus on intermittent demand*. PhD thesis, Luleå tekniska universitet, 2009.

[29] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295, 2016.

[30] Emil Willersjö Nyfelt. Comparison of the 1st and 2nd order Lee–Carter methods with the robust Hyndman–Ullah method for fitting and forecasting mortality rates, 2020.

[31] Peter Wilton. Spanish flu outdid WWI in number of lives claimed. *CMAJ: Canadian Medical Association Journal*, 148(11):2036, 1993.

Appendix A - R code

Sweden and France (WWII with the longer training set)

Here is presented the code used for the mortality forecasts for Sweden and France presented in Subsection 3.2.3. By changing extracted years, variable names and forecasting horizon, the similar code can be applied based any other available historical data.

```
library (demography) # Use demography library

######################################################################
# Fitting procedure (Import historical data from HMD) #
swe.rawData <- hmd.mx("SWE", "username","password", "Sweden")
swe.choice <- extract.years(swe.rawData, years=1872:1990)
swe.choice <- extract.ages(swe.choice, ages=0:90, combine.upper=FALSE)
swe.smooth <- smooth.demogdata(swe.choice, method="mspline", b=50, obs.var
   ="empirical")

fra.rawData <- hmd.mx("FRATNP","username","password","France")
fra.choice <- extract.years(fra.rawData, years=1872:1990) #swechoice
fra.choice <- extract.ages(fra.choice, ages=0:90,combine.upper=FALSE)
fra.smooth <- smooth.demogdata(fra.choice, method="mspline", b=50, obs.var
   ="empirical") # Monotony after age of 50 years

#extracting the years for smoothed training data
swe.smooth.train <- extract.years(swe.smooth, years=1872:1950)
fra.smooth.train <- extract.years(fra.smooth, years=1872:1950)
#extracting the years for testing data
swe.test <- extract.years(swe.smooth, years=1951:1990)
fra.test <- extract.years(fra.smooth, years=1951:1990)

######################################################################
#HU method
######################################################################
#fitting the HU method to the smoothed data
hu.swe.smooth.train <- fdm(swe.smooth.train, series="total", order=4)
hu.fra.smooth.train <- fdm(fra.smooth.train, series="total", order=6)
#forecasting from the training dataset
fhu.swe.smooth.train <- forecast(hu.swe.smooth.train, h=40) #h=horizon of
   40y
fhu.fra.smooth.train <- forecast(hu.fra.smooth.train, h=40)

# Check the % variation
summary(hu.swe.smooth.train)
hu.swe.percent = sum(hu.swe.smooth.train[["varprop"]])
summary(hu.fra.smooth.train)
hu.fra.percent = sum(hu.fra.smooth.train[["varprop"]])

######################################################################
#Weighted HU method
######################################################################
#fitting the wHU method to the smoothed data
```

```
whu.swe.smooth.train <- fdm(swe.smooth.train, series="total", method="
    classical", weight=TRUE, order=4)
whu.fra.smooth.train <- fdm(fra.smooth.train , series="total", method="
    classical", weight=TRUE, order=6)
#forecasting from the training dataset
fwhu.swe.smooth.train <- forecast(whu.swe.smooth.train, h=40)
fwhu.fra.smooth.train <- forecast(whu.fra.smooth.train, h=40)
#done to check the % variation
summary(whu.swe.smooth.train)
whu.swe.percent = sum(whu.swe.smooth.train[["varprop"]])
summary(whu.fra.smooth.train)
whu.fra.percent = sum(whu.fra.smooth.train[["varprop"]])

#comparisons of means
#comparison in HU
compare.demogdata(swe.test, series="total", fhu.swe.smooth.train)
compare.demogdata(fra.test, series="total", fhu.fra.smooth.train)
#comparison in wHU
compare.demogdata(swe.test, series="total", fwhu.swe.smooth.train)
compare.demogdata(fra.test, series="total", fwhu.fra.smooth.train)

########################################################################
########################################################################
swe.rawData_short <- hmd.mx("SWE", "username", "password", "Sweden")
swe.choice_short <- extract.years(swe.rawData_short, years=1900:1990)
swe.choice_short <- extract.ages(swe.choice_short, ages=0:90, combine.
    upper=FALSE)
swe.smooth_short <- smooth.demogdata(swe.choice_short, method="mspline", b
    =50, obs.var="empirical")

fra.rawData_short <- hmd.mx("FRATNP", "username", "password", "France")
fra.choice_short <- extract.years(fra.rawData_short, years=1900:1990)
fra.choice_short <- extract.ages(fra.choice_short, ages=0:90,combine.upper
    =FALSE)
fra.smooth_short <- smooth.demogdata(fra.choice_short, method="mspline", b
    =50, obs.var="empirical")

#extracting the years for smoothed training data
swe.smooth.train_short <- extract.years(swe.smooth_short, years=1924:1950)
fra.smooth.train_short <- extract.years(fra.smooth_short, years=1924:1950)
#extracting the years for testing data
swe.test_short <- extract.years(swe.smooth_short, years=1951:1990)
fra.test_short <- extract.years(fra.smooth_short, years=1951:1990)

########################################################################
#HU method
########################################################################
#fitting the HU method to the smoothed data
hu.swe.smooth.train_short <- fdm(swe.smooth.train_short, series="total",
    order=4)
hu.fra.smooth.train_short <- fdm(fra.smooth.train_short, series="total",
    order=6)
```

```
#forecasting from the training dataset
fhu.swe.smooth.train_short <- forecast(hu.swe.smooth.train_short, h=40)
fhu.fra.smooth.train_short <- forecast(hu.fra.smooth.train_short, h=40)

# Check the % variation
summary(hu.swe.smooth.train_short)
hu.swe.percent_short = sum(hu.swe.smooth.train_short[["varprop"]])
summary(hu.fra.smooth.train_short)
hu.fra.percent_short = sum(hu.fra.smooth.train_short[["varprop"]])

##########################################################################
#Weighted HU method
##########################################################################
#fitting the wHU method to the smoothed data
whu.swe.smooth.train_short <- fdm(swe.smooth.train_short, series="total",
    method="classical", weight=TRUE, order=4)
whu.fra.smooth.train_short <- fdm(fra.smooth.train_short , series="total",
     method="classical", weight=TRUE, order=6)
#forecasting from the training dataset
fwhu.swe.smooth.train_short <- forecast(whu.swe.smooth.train_short, h=40)
fwhu.fra.smooth.train_short <- forecast(whu.fra.smooth.train_short, h=40)
#done to check the % variation
summary(whu.swe.smooth.train_short)
whu.swe.percent_short = sum(whu.swe.smooth.train_short[["varprop"]])
summary(whu.fra.smooth.train_short)
whu.fra.percent_short = sum(whu.fra.smooth.train_short[["varprop"]])

##########################################################################
#Plotting in-sample forecast (WWII) - Sweden
##########################################################################
par(mfrow = c(2, 2),
    oma = c(2, 1, 1, 0),
    mar = c(1, 2, 2, 1),
    mgp = c(3, 1, 0))

#year1955
plot(swe.choice, years=1955, series="total", ylim=c(-10,-1), col = "blue",
     lty=1, lwd=2, main="")
lines(fhu.swe.smooth.train, years=1955, col="red", lty=1, lwd=1, pch=1,
    type = "b")
lines(fwhu.swe.smooth.train, years=1955, col="black", lty=5, lwd=2)
lines(fhu.swe.smooth.train_short, years=1955, col="magenta4", lty=1, lwd
    =1, pch=1, type = "b")
lines(fwhu.swe.smooth.train_short, years=1955, col="azure4", lty=5, lwd=2)
text(x = 45, y = -1.5, label = "1955", col = "black", font = 2, cex = 1.5)
legend("bottomright", legend=c("Actual Data","HU Method", "wHU Method"),
       cex=1.00, bty = "n", bg='white', col=c(12,2,1,1), lty=c(1,1,5,1),
    lwd=1.2,
       pt.cex = 1.2, pch = c(NA, 1, NA), x.intersp=0.4, y.intersp=0.5,
       text.width=17.00, text.font = 2)
text(x = 75, y = -7.2, label = "Training set: 1872-1950", col = "black",
    font = 2, cex = 1.0)
```

```
legend("topleft", legend=c("HU Method", "wHU Method"), inset = 0.03,
        cex=1.00, bty = "n", bg='white', col=c("#8B008B","#838B8B",1,1), lty=c(1,5,5,1), lwd=1.2,
        pt.cex = 1.2, pch = c(1, NA, NA), x.intersp=0.4, y.intersp=0.5,
        text.width=17.00, text.font = 2)
text(x = 12, y = -1.4, label = "Training set: 1924-1950", col = "black", font = 2, cex = 1.0)
grid(lwd=1)

#year1960
plot(swe.choice, years=1960, series="total", ylim=c(-10,-1), col = "blue",
     lty=1, lwd=2, main="")
lines(fhu.swe.smooth.train, years=1960, col="red", lty=1, lwd=1, pch=1, type = "b")
lines(fwhu.swe.smooth.train, years=1960, col="black", lty=5, lwd=2)
lines(fhu.swe.smooth.train_short, years=1960, col="magenta4", lty=1, lwd=1, pch=1, type = "b")
lines(fwhu.swe.smooth.train_short, years=1960, col="azure4", lty=5, lwd=2)
text(x = 45, y = -1.5, label = "1960", col = "black", font = 2, cex = 1.5)
legend("bottomright", legend=c("Actual Data","HU Method", "wHU Method"),
        cex=1.00, bty = "n", bg='white', col=c(12,2,1,1), lty=c(1,1,5,1), lwd=1.2,
        pt.cex = 1.2, pch = c(NA, 1, NA), x.intersp=0.4, y.intersp=0.5,
        text.width=17.00, text.font = 2)
text(x = 75, y = -7.2, label = "Training set: 1872-1950", col = "black", font = 2, cex = 1.0)
legend("topleft", legend=c("HU Method", "wHU Method"), inset = 0.03,
        cex=1.00, bty = "n", bg='white', col=c("#8B008B","#838B8B",1,1), lty=c(1,5,5,1), lwd=1.2,
        pt.cex = 1.2, pch = c(1, NA, NA), x.intersp=0.4, y.intersp=0.5,
        text.width=17.00, text.font = 2)
text(x = 12, y = -1.4, label = "Training set: 1924-1950", col = "black", font = 2, cex = 1.0)
grid(lwd=1)

#year1965
plot(swe.choice, years=1965, series="total", ylim=c(-10,-1), col = "blue",
     lty=1, lwd=2, main="")
lines(fhu.swe.smooth.train, years=1965, col="red", lty=1, lwd=1, pch=1, type = "b")
lines(fwhu.swe.smooth.train, years=1965, col="black", lty=5, lwd=2)
lines(fhu.swe.smooth.train_short, years=1965, col="magenta4", lty=1, lwd=1, pch=1, type = "b")
lines(fwhu.swe.smooth.train_short, years=1965, col="azure4", lty=5, lwd=2)
text(x = 45, y = -1.5, label = "1965", col = "black", font = 2, cex = 1.5)
legend("bottomright", legend=c("Actual Data","HU Method", "wHU Method"),
        cex=1.00, bty = "n", bg='white', col=c(12,2,1,1), lty=c(1,1,5,1), lwd=1.2,
        pt.cex = 1.2, pch = c(NA, 1, NA), x.intersp=0.4, y.intersp=0.5,
        text.width=17.00, text.font = 2)
text(x = 75, y = -7.2, label = "Training set: 1872-1950", col = "black", font = 2, cex = 1.0)
```

```
legend("topleft", legend=c("HU Method", "wHU Method"), inset = 0.03,
       cex=1.01, bty = "n", bg='white', col=c("#8B008B","#838B8B",1,1),
    lty=c(1,5,5,1), lwd=1.2,
       pt.cex = 1.2, pch = c(1, NA, NA), x.intersp=0.4, y.intersp=0.5,
       text.width=17.00, text.font = 2)
text(x = 12, y = -1.4, label = "Training set: 1924-1950", col = "black",
    font = 2, cex = 1.0)
grid(lwd=1)

#year1970
plot(swe.choice, years=1970, series="total", ylim=c(-10,-1), col = "blue",
     lty=1, lwd=2, main="")
lines(fhu.swe.smooth.train, years=1970, col="red", lty=1, lwd=1, pch=1,
    type = "b")
lines(fwhu.swe.smooth.train, years=1970, col="black", lty=5, lwd=2)
lines(fhu.swe.smooth.train_short, years=1970, col="magenta4", lty=1, lwd
    =1, pch=1, type = "b")
lines(fwhu.swe.smooth.train_short, years=1970, col="azure4", lty=5, lwd=2)
text(x = 45, y = -1.5, label = "1970", col = "black", font = 2, cex = 1.5)
legend("bottomright", legend=c("Actual Data","HU Method", "wHU Method"),
       cex=1.00, bty = "n", bg='white', col=c(12,2,1,1), lty=c(1,1,5,1),
    lwd=1.2,
       pt.cex = 1.2, pch = c(NA, 1, NA), x.intersp=0.4, y.intersp=0.5,
       text.width=17.00, text.font = 2)
text(x = 75, y = -7.2, label = "Training set: 1872-1950", col = "black",
    font = 2, cex = 1.0)
legend("topleft", legend=c("HU Method", "wHU Method"), inset = 0.03,
       cex=1.01, bty = "n", bg='white', col=c("#8B008B","#838B8B",1,1),
    lty=c(1,5,5,1), lwd=1.2,
       pt.cex = 1.2, pch = c(1, NA, NA), x.intersp=0.4, y.intersp=0.5,
       text.width=17.00, text.font = 2)
text(x = 12, y = -1.4, label = "Training set: 1924-1950", col = "black",
    font = 2, cex = 1.0)
grid(lwd=1)
```

Sweden (Out-of-Sample)

Here is presented the code used for out-of-sample mortality forecasting of Sweden. Apply the same procedure for other countries.

```
library (demography) # Use demography library
##########################################################################
# Fitting procedure (Import historical data from HMD) #
swe.rawData <- hmd.mx("SWE", "username","password", "Sweden")
swe.choice <- extract.years(swe.rawData, years=1951:2020)
swe.choice <- extract.ages(swe.choice, ages=0:90, combine.upper=FALSE)
swe.smooth <- smooth.demogdata(swe.choice, method="mspline", b=50, obs.var="empirical")
##########################################################################
# Out-of-sample forecast - Sweden
##########################################################################
#fitting the wHU to the smoothed dataset
swe.fit <- fdm(swe.smooth, series="total", method="classical",
               weight=TRUE, beta=0.1, order=5)

#forecasting from the fitted model
fcast.sweden <- forecast(swe.fit, h=30, level=95, method="arima")
#extracting historical data
swe2020=extract.years(swe.smooth, years=1971:2020)

#plotting forecast with historical data
plot(fcast.sweden, main="", xlim=c(0,90), ylim=c(-11,0), las=1)
axis(1, at = seq(0, 90, by = 5))
lines(swe2020, series="total", lty=3, lwd=2)
legend("bottomright",  legend=unique(fcast.sweden$year),
       col=rainbow(length(fcast.sweden$year)*1.25), ncol=5 , pch=NA,
       bty ="n", bg ='white', x.intersp=0.4, y.intersp=0.30,
       text.width=5.00, pt.cex=1, cex=0.80, text.font=2, lty=1, lwd=3)
text(x = 65, y = -7.8, label = "Forecast Years", col = "black", font = 2, cex = 1.0)
legend("topleft", inset=0.04, legend=unique(swe2020$year),
       col=rainbow(length(fcast.sweden$year)*1.25), ncol=5, pch=NA,
       bty ="n", bg ='white', x.intersp=0.4, y.intersp=0.30,
       text.width=5.00, pt.cex=1, cex=0.80, text.font=2, lty=3, lwd=4)
text(x = 25, y = -0.3, label = "Actual Years", col = "black", font = 2, cex = 1.0)
grid(lwd=1)

#plot the basis functions and associated coefficients
plot(fcast.sweden, plot.type="component")

#checking the % variation
summary(swe.fit)
swe.var=sum(swe.fit[["varprop"]])
models(fcast.sweden)
```

Appendix B - The basis functions and associated coe□cients

By calling the function `plot(fcast.sweden, plot.type="component")` where `fcast.sweden` is a list containing the forecasting results, the plots of the basis functions and associated coe□cients are obtained. The same procedure is applied for each of the tested countries.

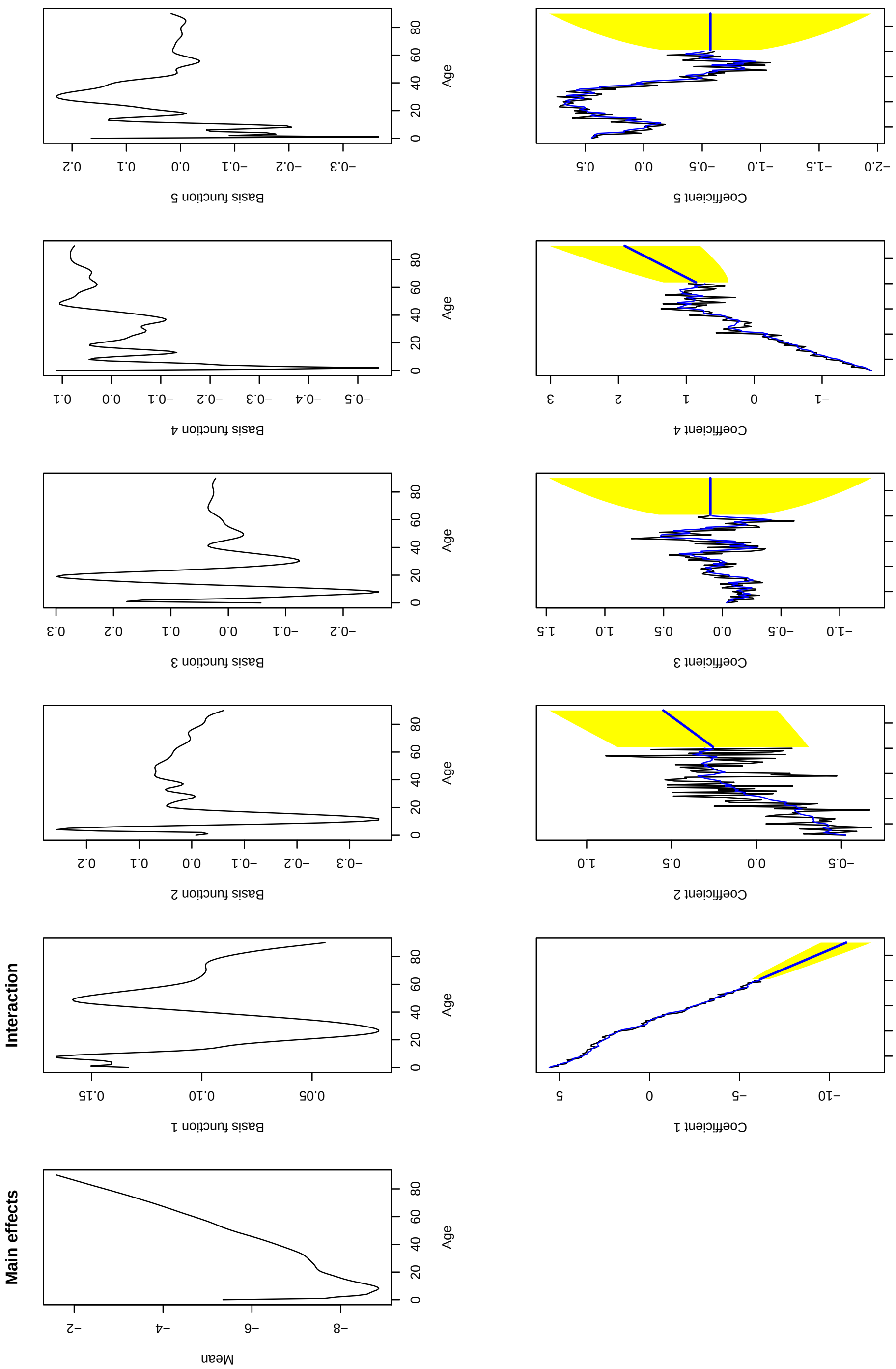

Figure 1: The basis functions and associated coe□cients for Sweden.

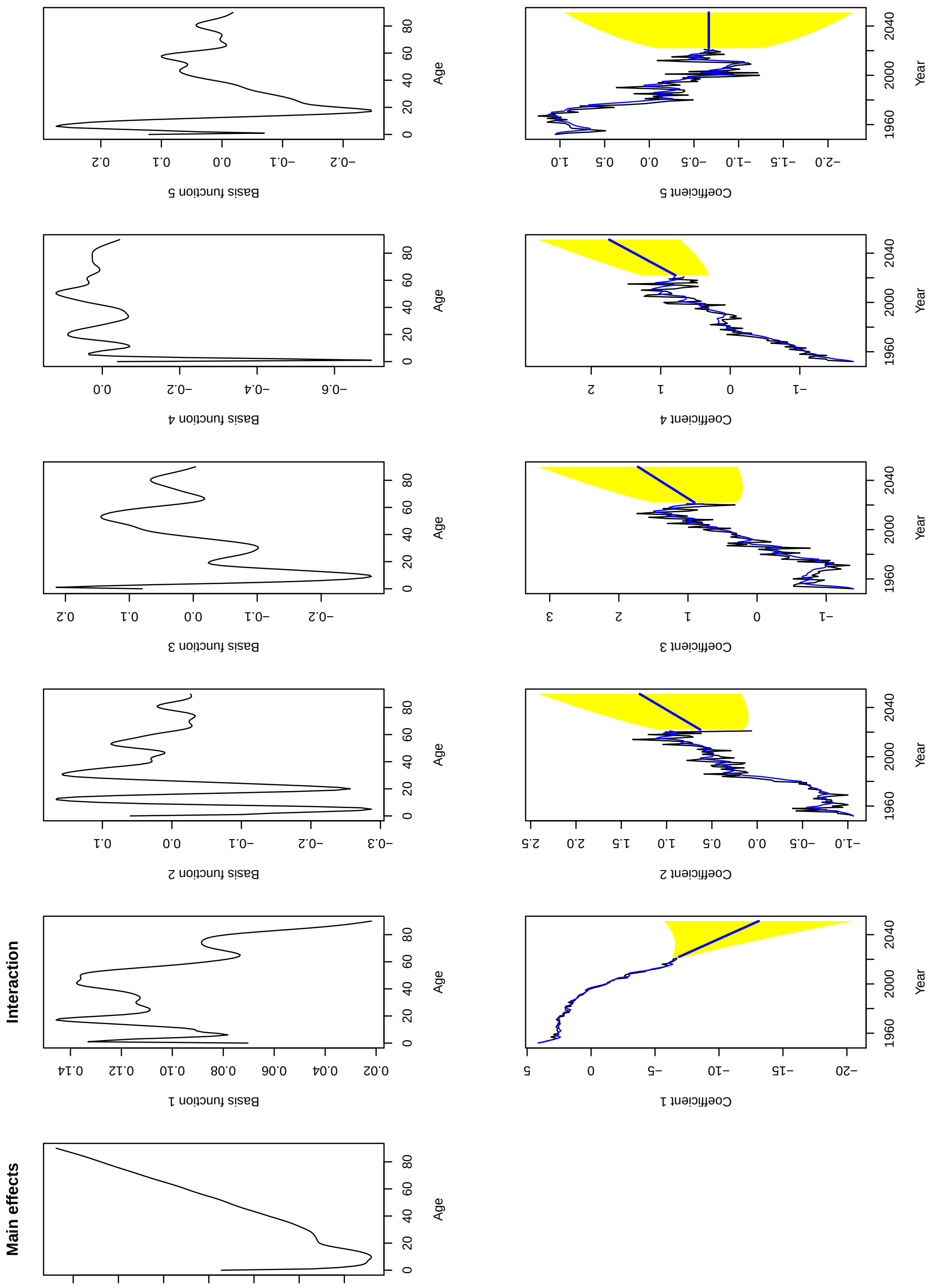

Figure 2: The basis functions and associated coe□cients for Denmark.

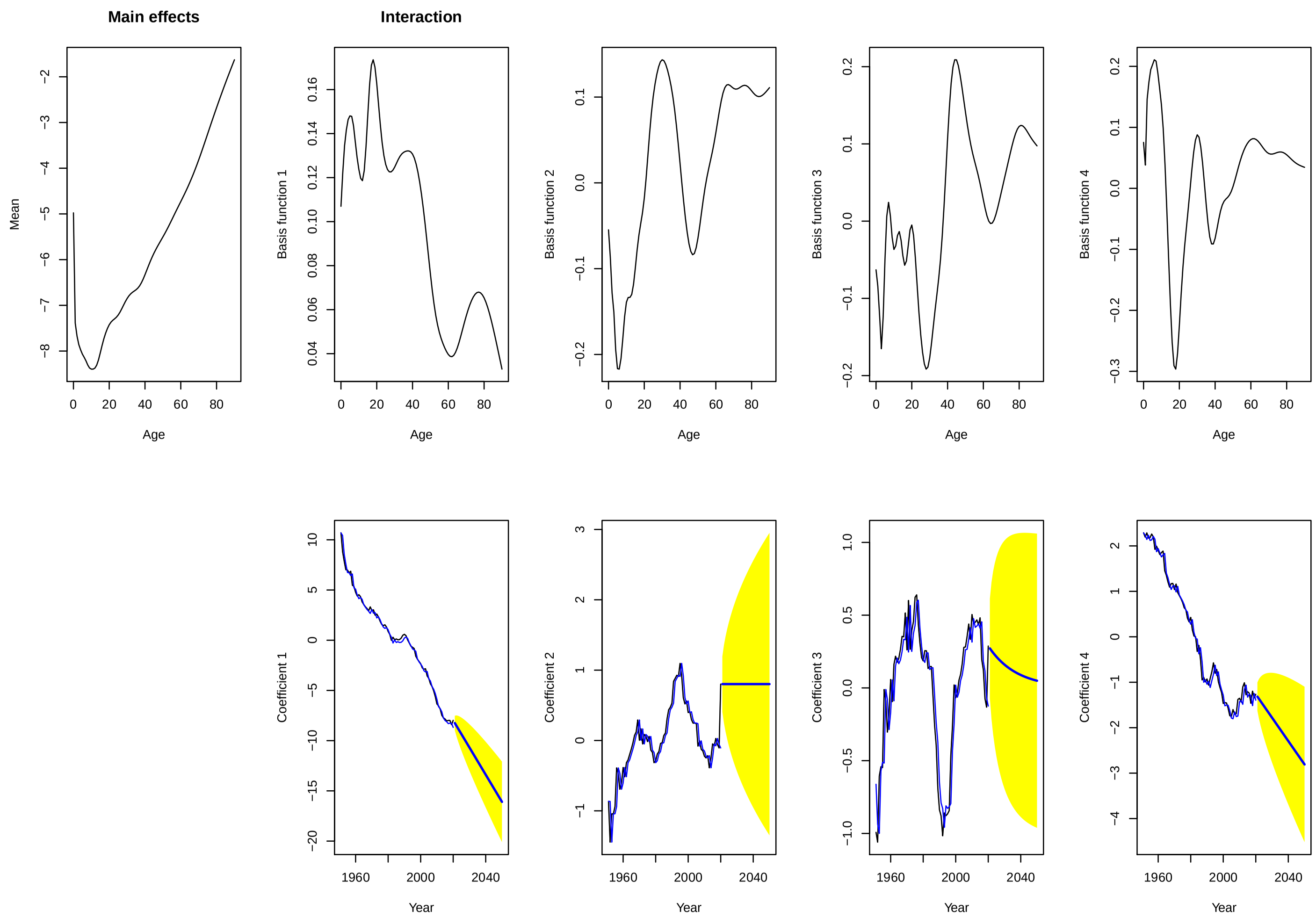

Figure 3: The basis functions and associated coe□cients for Spain.

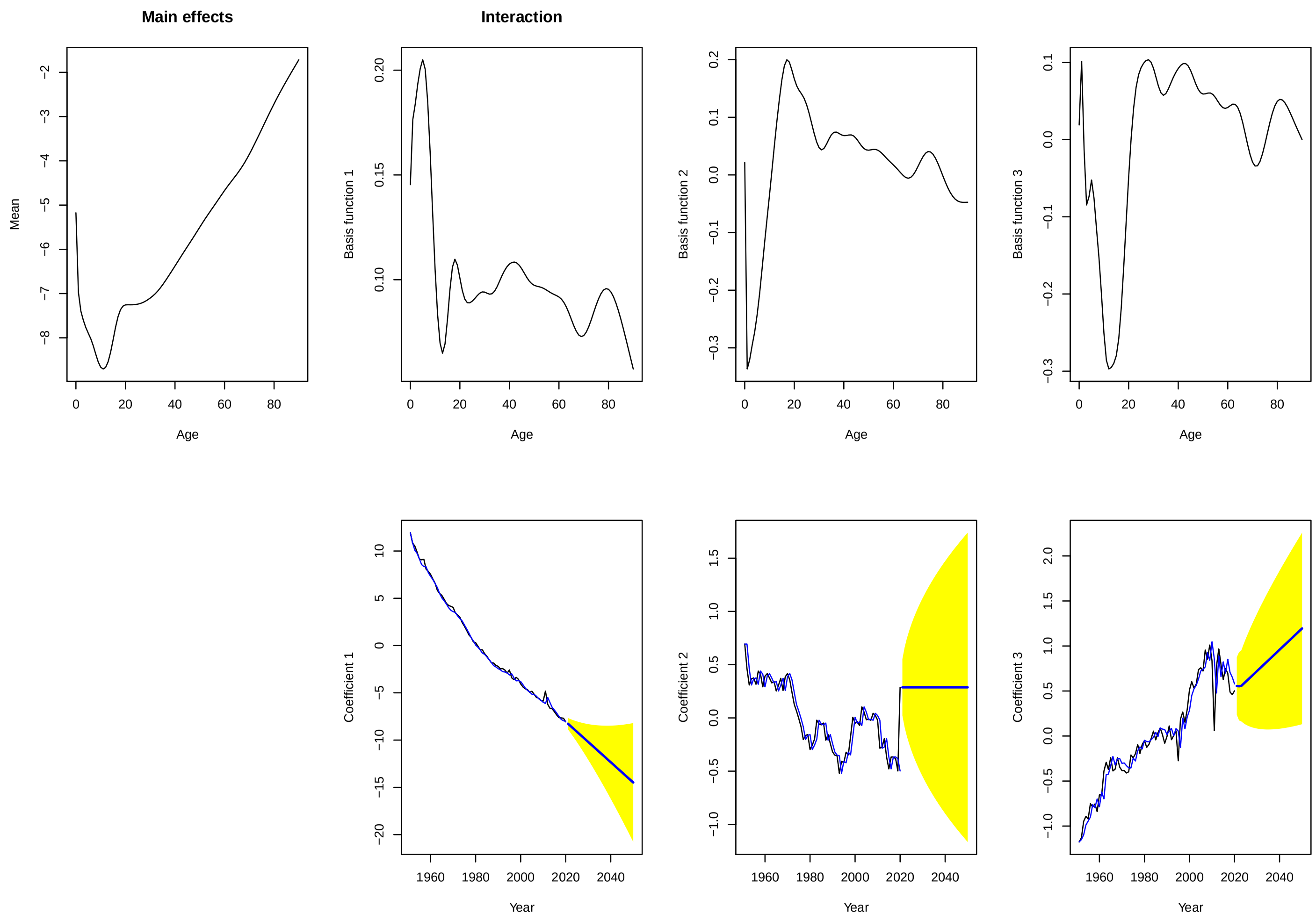

Figure 4: The basis functions and associated coe□cients for Japan.

Appendix C - ARIMA forecast coefficients

Here is presented the output from the R software regarding out-of-sample forecasting for Sweden. The same procedure is applied for all other countries by considering other lists instead of `fcast.sweden`.

```
> models(fcast.sweden)

-- Coefficient 1 --
Series: xx[, i]
ARIMA(0,1,1) with drift

Coefficients:
          ma1     drift
      -0.6349   -0.1830
s.e.   0.1647    0.0132

sigma^2 estimated as 0.05995:  log likelihood=0.18
AIC=5.63   AICc=6.17   BIC=11.31

-- Coefficient 2 --
Series: xx[, i]
ARIMA(0,1,1)

Coefficients:
          ma1
      -0.8742
s.e.   0.0839

sigma^2 estimated as 0.1079:  log likelihood=-15.19
AIC=34.39   AICc=34.65   BIC=38.17

-- Coefficient 3 --
Series: xx[, i]
ARIMA(1,0,0) with zero mean

Coefficients:
         ar1
      0.4877
s.e.  0.1208

sigma^2 estimated as 0.0544:  log likelihood=2.21
AIC=-0.42   AICc=-0.16   BIC=3.41

-- Coefficient 4 --
Series: xx[, i]
ARIMA(0,1,1) with drift

Coefficients:
          ma1    drift
      -0.6531   0.0263
s.e.   0.1158   0.0143
```

```
sigma^2 estimated as 0.08024:  log likelihood=-6.98
AIC=19.95   AICc=20.49   BIC=25.63

-- Coefficient 5 --
Series: xx[, i]
ARIMA(0,1,1)

Coefficients:
          ma1
      -0.4181
s.e.   0.1230

sigma^2 estimated as 0.04636:  log likelihood=6.13
AIC=-8.26   AICc=-8   BIC=-4.47
```

Appendix D - Thesis Objectives

Objective 1: Knowledge and understanding

I believe that this thesis serves as an insightful representation of the knowledge and understanding of the presented topic. It specializes in the fields of demography, actuarial mathematics and time series analysis, with the problem of forecasting mortality rates using HU method taking into account Covid-19 mortalities. A lot of time has been spent in the process of selecting, evaluating and interpreting the literature that contributed to this research project. By explaining all the significant concepts in the Chapter 2 before implementing the real-life problem in the Chapter 3, the reader is offered the chance to properly understand the reasoning and purpose of this thesis.

Objective 2: Methodological knowledge

As an author, I gave my best to methodically combine the knowledge obtained from the courses of actuarial mathematics, time series analysis, matrix analysis, applied linear algebra and other fields of study that contributed to the covered topic. Additionally, I read and analyzed numerous relevant papers and reports and extracted the most important concepts and definitions, while referring to the literature so the reader can find out more about the certain topic. In that way, individuals with little or no background in this field will be able to understand the content appropriately.

Objective 3: Critically and systematically integrate knowledge

During the entire time of writing this thesis, various books, articles and scientific reports had to be evaluated and critically interpreted before integrating the certain knowledge into the content. This is proven by the list of more than 30 references in the part 4.2. Forecasting life expectancies was not among the components of the initial idea, but my supervisor Milica Rancic appreciated my idea of making the most of the package *demography*. Due to their close connection with the mortality rates, the tables of the life expectancies might offer additional conclusions about the efficiency of the implemented HU and wHU methods.

Objective 4: Independently and creatively identify and carry out advanced tasks

While investigating the topic, I came across several interesting paths that this report can be directed. This thesis has been written by the student, but the important role of my supervisor needs to be mentioned. Keeping constant and active contact on a biweekly basis, we organized the structure and aims of the report. I am proud that we successfully overcame all the obstacles that appeared during the process. For example, the out-of-sample forecast was highly dependable on the available historical data, so some initial ideas could not be realized desirably. However, I believe that my work combined with the supervisor's support resulted in very insightful research. To not allow the possibility of breaking the deadlines, the report was written almost daily, followed by an extensive investigation regarding the new challenges.

Objective 5: Present and discuss conclusions and knowledge

The structure of this thesis follows the natural and logical path. First, the broad topic of mortality rates was introduced, together with the summary of the evolution of mortality forecasting throughout the history. Then, all the crucial concepts and methodology were presented, such are ARIMA models, fPCA, Weighted Penalized Regression Splines and HU and wHU models. However, to not diverge too far from the main point of interest, we decided to keep the theoretical background concise. The large portion of this report consists of the implementation of the forecasting models, which offers certain conclusions presented after every graph or table of results.

Objective 6: Scientific, social and ethical aspects

I believe that this thesis contains all of the main properties of the scientific report. The collected information are approached critically, while all used literature is cited in a clear way. The Human mortality database served as the primary source of the mortality data, which is already stated as being the basis of the performed implementation. The obtained results are thoroughly examined and explained.

From the social and ethical aspects, the forecasts obtained by applying HU and wHU methods considering the outlying years caused by Covid-19 pandemics are presented. Regarding the topic of human mortalities, this thesis is intended to be one of the many different researches that contributed in raising the knowledge of the actual human lives by forecasting mortalities as accurately as possible.