LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
DEPARTMENT OF STATISTICS

Modern approaches in mortality modeling considering data enrichment and the impact of COVID-19

MASTER THESIS

Author:
Asmik Nalmpatian

Supervisor: Prof. Dr. Christian Heumann
Advisor: Dr. Stefan Pilz, Munich Re

Date: September 20, 2021

Acknowledgements

First of all, I want to thank my supervisor Prof. Dr. Christian Heumann for his willingness to accompany and supervise me on the way to this thesis, to support me with valuable advice, to correct and direct my ideas.

I am indebted to Dr. Stefan Pilz for his incredibly professional help with the extensive literature research and fruitful comments, for his motivation and guidance in challenging moments.

Furthermore, I would like to thank Dr. Andreas Bayerstadler and Dr. Stefan Pilz for the opportunity to write my thesis in cooperation with Munich Re and to explore a new field of application for my statistical expertise.

Finally, I thank my family and the people close to me for their love and unconditional support during the whole process of this thesis and throughout my studies.

Abstract

During the last two centuries, substantial increase in life expectancy was observed. Even though past developments suggest that the mortality will further decrease in the future, the uncertainty and instablity about the evolution is highly intensified due to the ongoing COVID-19 pandemic. Especially for both pension scheme and life insurers it is therefore of essential interest to predict the mortality of its members or policyholders with reliable accuracy.

The aim of this thesis is twofold. Firstly, to develop an appropriate enrichment method to map the mortality data from categorized into metric age scale. Based on this, different approaches to both fit and forecast human mortality in Finland, Germany, Italy, the Netherlands and the United States are applied. The special focus lies on improving the state-of-the-art stochastic models with machine learning techniques and generalizing them to a multipopulational model as well as enriching the analysis with further socio-economic and health-risk variables.

The second part is devoted to meeting trend forecasts of mortality rates and life expectancies as a measure of longevity. To enable this, different - from mild to severe - scenarios for the future are set up, assessed for plausibility and discussed. This is helpful for preparedness, planning and informed decision making despite the current situation surrounded by biological and epidemiological ambiguities.

The thesis ends by providing numerical calculations for age-specific life expectancies for the years 2022 and 2023 with specific consideration of the COVID-19 impact.

Contents

1 Introduction

The developed world has experienced a substantial increase in life expectancy during the last decades. Data for the USA shows, for example, that the number of deaths of 80 years-old men per thousand decreased by 56,7 % from 1933 to 2019 (from 127 to 55 deaths per thousand) (HMD, 2021). Indeed, this decline in mortality rates is associated with social and health advances and makes it particularly interesting for governments, private pension companies, and life insurers to build models and estimates to make projections for the future. The mortality rate improvements must therefore be captured and forecasted adequately, in order to provide a solid foundation for decision making.

Quantifying the uncertainty about mortality trends is of high relevance, especially after being intensified by COVID-19. Due to the pandemic, mortality rates have been discussed for a broader audience. Generally, mortality rate projections consist of two parts. Firstly, estimating a mortality model based on historical data, which captures the mortality rates by taking into account three key components: age, time period and birth cohort. Secondly, forecasting the time dependent parameters or trends from the estimated model into the future. Hereby, it is important to properly predict the future extent of population ageing, so that the sustainability of pension schemes and social security systems are ensured.

This thesis aims to use data and statistical models to obtain reliable estimates and forecasts as adequate models help insurance companies to meet their liabilities and offer products to customers. Both pension scheme and life insurers aim to predict the mortality of its members or policyholders with reliable accuracy. On the one hand, this will help a pension scheme to handle liabilities in an effective fashion and to make informed decisions. On the other hand, the life insurer would then be able to set appropriate premiums, so that they are high enough to cover the policyholders and, at the same time, as low as possible to remain competitive and worthwhile for the clients. Although some factors are known to influence human mortality, it is not a deterministic process by nature. Therefore, an effort is made to estimate how the random components of mortality rates will develop in the future for different population groups based on historical data. This leads to the first research question:

How can state-of-the-art mortality models be improved in terms of fit and forecast?

The most widely used Lee Carter model will serve as a benchmark mainly due to its robustness, simplicity and explanation, taking into account the age and period effect (Lee and Carter, 1992). Many variants, extensions and alternatives to the Lee Carter model exist, such as the CBD model (Cairns et al., 2006) and the so-called Quadratic CBD model with additional cohort effects (M7) (Cairns et al., 2009) as well as the Age-Period-Cohort model (APC) (Clayton and Schifflers, 1987, Hobcraft et al., 1982, Holford, 1983, Currie, 2006). These four stochastic modelling approaches will be compared separately for each subpopulation by gender and country analyses, as they do not allow for joint estimations in their original versions. Each subpopulation contains the male or female population of a given country. The identifiability problems inherent in some of these models are regulated by constraints. The following countries will be examined in detail: Germany, Finland, the Netherlands and Italy represent Europe and the United States represent North

America. We compare different approaches to human mortality adjustment and prediction to identify defects in parameter estimation and to further develop and propose effective methods. For this, we use the data from Human Mortality Database (HMD, 2021). This dataset provides aggregate data that includes country, year, age, sex, as well as population size, number of deaths, and mortality rates as variables. As the interest of the modelling approaches lies in the precise analysis of the age, cohort and period effects, a methodology has been developed to map the mortality rates from grouped to metric age scale for recent years, so that the data can be enriched properly. The data on annual and metric age level is only available until 2019 or even before, depending on the provided information by the different countries. Another dataset, also from Human Mortality Database, provides this missing information up to the current date but with grouped age information only. In the second step, the Lee Carter model estimation will be revisited to be adequately improved by tree-based machine learning techniques. The aim is to detect the weaknesses of the Lee Carter model and to improve them. The time-dependent components (i.e. period and cohort) are forecasted using ARIMA models or random walk with drift, taking into account the structural breaks indicating unexpected changes over time and dated in the period effects (Bai and Perron, 1998, Zeileis et al., 2003). However, all these models predict the future mortality of a single population without considering the impact of neighbour countries, their potential dependencies on the geopolitical, socio-economic or any other common criteria. Therefore, a multipopulational approach for simultaneous modelling of mortality in two (or more) related populations will be introduced to enable borrowing strength from similar data. This will illustrate how the classical APC model can be further developed in the semi-parametric framework of Generalized Additive Models (GAM) and improved in the sense of fitting and capturing the uncertainty (Clements et al., 2005). With this approach there is an identification problem, too. The three components of the temporal structures age, period and cohort depend on each other so that each of them is a linear combination of the other two (i.e. cohort is calculated by the difference between period and age). Thus, the assignment of the interrelations is not unique. By fitting a bivariate tensor product spline between age and period, we can avoid these identification problems without running into limitations (Wood, 2017). This helps to overcome the identification challenge of understanding the interrelations of the three main components age, period and cohort without meeting further assumptions and constraints. In addition, it is relevant to consider the variability in mortality improvements within each country, keeping the socio-economic and health-related differences in mind. This poses an additional challenge on the accurate design of pension systems and the management of longevity risk in pension funds and annuity portfolios. Therefore, apart from jointly analysing the age period and cohort effects for different gender groups and countries, further variables that are potentially associated with the mortality rates will be included into the model. For this a data enrichment procedure will be taken to integrate additional variables on annual country level from various open data sources, such as the World Health Organization (WHO, 2021), OECD (OECD, 2021) and the United Nations (UIS, 2021). The contribution of the inclusion of these variables to model performance will be discussed subsequently.

The coronavirus disease (COVID-19), first broken out in the Chinese city Wuhan in December 2019, has spread to nearly every country (Mohan and Nambiar, 2020).

At the time of preparing this thesis, there have been reported nearly 177 million cases and 4 million deaths worldwide (Johns Hopkins, 2021). As this is an ongoing event and the impact of COVID-19 on the frailty of survivals, the further development and the long-term effects are not clear yet, the mortality rates of the year 2020 require special handling in terms of forecasting. This is also an incentive to get up-to-date data and thus, the investigation of the second research question can be formulated as follows:

How will mortality develop in the future in different countries, considering the COVID-19 impact?

On the basis of GAM, several forecasting scenarios will be developed, applied and discussed with regard to each underlying assumption. These scenarios should be seen as possible examples and the results should be interpreted with caution, particularly because of the uncertainty regarding virus variants and vaccinations, the consequences are not sufficiently foreseeable at the time of writing this thesis. Special focus lies on the consideration of the excess mortality. If one does not incorporate it, the prediction of the following years will be biased due to censoring (corona death cases). Based on this information, life expectancy tables are provided for the next two years.

The thesis is organized as follows. In Section 2, the database will be introduced and described within the framework of an exploratory data analysis. In Section 3, the enrichment strategies in terms of the mapping of mortality rates from grouped into metric age scale and the integration of additional covariates will be explained. Section 4 gives a theoretical overview of all applied methods from Lee Carter and variants, machine learning algorithms, GAMs via evaluation metrics to forecasting approaches. In Section 5, the results for all countries and all models will be introduced and compared. Discussion and conclusion are then given in Sections 6 and 7, respectively.

2 Database

The Human Mortality Database (HMD, 2021) provides data on mortality rates $\mu_{a,t,s}$ as well as the number of deaths $D_{a,t,s}$ during calendar year t aged a in years at death for a given subpopulation s. It also covers the population size $E_{a,t,s}$ of these people on January 1^{st} in year t, for each subpopulation s, comprising the country and gender. Data are included for both women and men from age 0 to 100+ years in the original version of the dataset. The mortality rates are calculated as follows:

$$\mu_{a,t,s} = \frac{D_{a,t,s}}{E_{a,t,s}} \tag{1}$$

The map in figure 1 is supposed to give an overview of the countries considered in this thesis. The motivation for selecting these countries is primarily their geographic context, as well as the relatively different course of the pandemic, in order to be able to crystallize the contrast in modeling: Finland, Germany, Italy and the Netherlands represent Europe, whereas the United States represent North America.

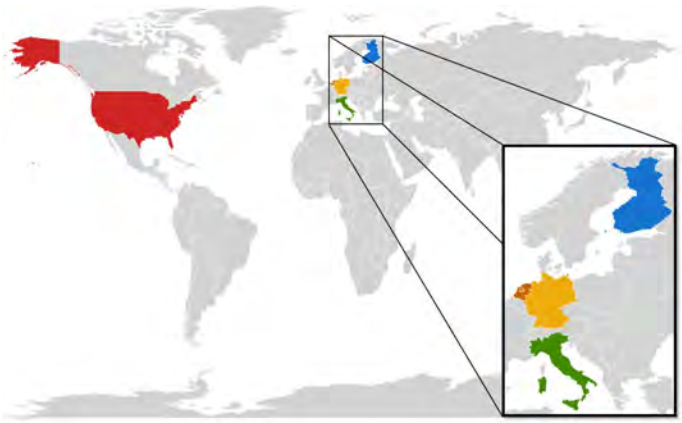


Figure 1: World map with the countries relevant for the present thesis

Due to the small number of observations in the higher age groups, the range of values for age is limited upward to 90 for further analysis. Both male and female population were used. Table 1 provides an overview of the availability of data for different countries.

Country	Available years
Finland	1850-2019
Germany	1990-2017
Italy	1872-2018
Netherlands	1850-2019
United States	1933-2019

Table 1: Temporal availability of the data in HMD (2021)

To enrich the data up to 2020 another dataset from HMD (2021) was used, called Short Term Mortality Fluctuations (STMF, 2021). This dataset provides information about weekly death counts and death rates reported for age groups $[0, 14], [15, 64], [65, 74], [75, 84], [85+]$ corresponding to $[l, u]$ instead of individual ages a, including the lower bound l and the upper bound u. The mapping methodology from weekly categorized mortality data into annual metric age scale will be described in detail in Section 3. The data availability of this dataset is given in Table 2:

Country	Available years
Finland	1990-2020
Germany	2016-2020
Italy	2011-2020
Netherlands	1995-2020
United States	2015-2020

Table 2: Temporal availability of the data in STMF (2021)

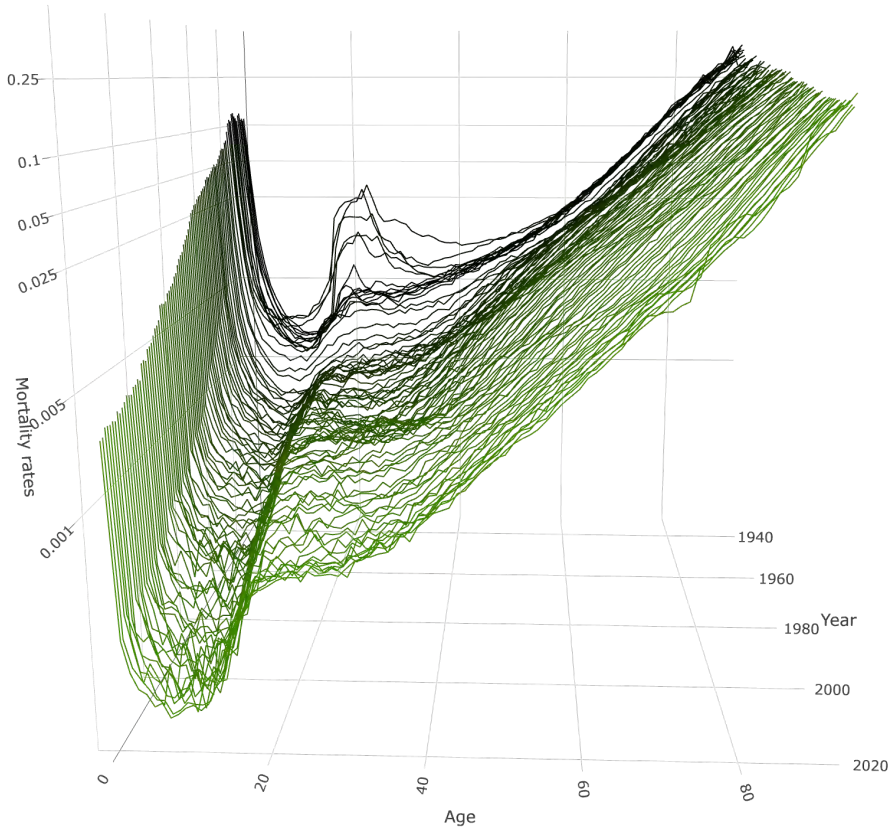


Figure 2: Evolution of mortality rates for Italian males aged 0–90 for the years 1933–2020. The data for year 2020 is enriched by the methodology that is developed and described in Section 3. Vertical axis spacing represents mortality rates with ticks on a logarithmic scale

Figures 2 and 3 provide an initial overview of the development of mortality rates. They illustrate the course of mortality rates simultaneously across years and ages for Italian and US male population in three dimensional graphs. The mortality surface shows that during the first part of 20^{th} century, infant mortality was high but improved and stabilized in the course of time clearly, probably due to medical, technological and hygienic advance as well as the improvement in living conditions (Wegman, 2001). This holds for all other ages, even though the steepest decrease is referred to the infant mortality. Up to an age of about 50 years, the mortality surface reveals some humps along the years, with mortality gradually increasing as individuals get older. These humps tend to occur in certain years, especially from 1918 to 1944, potentially linked to the influenza epidemy and to consequences of World War I or World War II (Wilson et al., 2014, Horiuchi, 1983). Apart from these humps the course over the ages remains quite similar over the years decreasing until the age of approximately 12, arriving at the steepest point there and then increasing slightly with growing age. For other male subpopulations, the corresponding figures are included in the appendix.

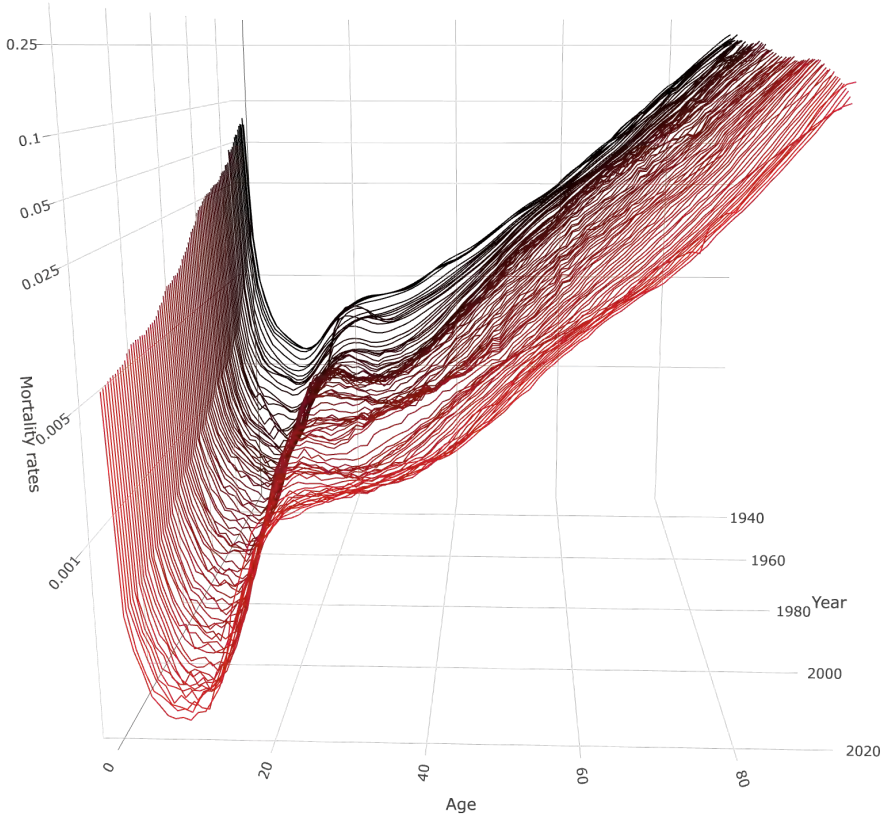


Figure 3: Evolution of mortality rates for US males aged 0–90 for the years 1933–2020. The data for year 2020 is enriched by the methodology that is developed and described in Section 3. Vertical axis spacing represents mortality rates with ticks on a logarithmic scale

After enriching the first dataset according to the mapping methodology described and explained in Section 3, the following years will be used for the further analysis: For single-population modelling approaches, data from 1950 to 2020 in case of Finland, the Netherlands, Italy and the United States and from 1990 to 2020 in case of Germany, due to data inconsistencies before re-unification. The cut in 1950 is proposed by Lee and Miller (2001) and justified by too many fluctuations before this year. For coherent modelling with GAM a joint coverage of years is required so that the available years for all countries will be restricted to 1990-2020, which is probably still sufficient for capturing more current effects on mortality rates and to project them into the future.

The heatmaps in figure 4 show the decrease in mortality rates in a more precise way for the relevant years and ages in the United States as an example. This is inspired by the so-called Lexis diagrams, i.e. a two-dimensional table or graph depicting age groups and periods in horizontal and vertical direction, respectively (Carstensen, 2007). For the other countries, considered in this thesis, the course remains overall similar. The females are depicted on the left-hand side and the males on the right-hand side. The color scheme indicates low (dark blue) to high (dark red) mortality rates, year and age are displayed on the $x-$ and $y-$axis, respectively. The slightly increasing, upward diagonal structure of the colors shows the typical course of mortality rates improvements for the last years. It also appears that females have on average lower mortality rates than males. The color pattern suggests that the

differences are more likely to affect the higher age categories, around 30 and above. These differences tend to be much larger in earlier years and seem to narrow over years. In order to delve deeper into noticeable deviations from the general trend, one should look at the bottom row depicting the mortality rates for age 0 and the right column representing the year 2020. The infant mortality has experienced a tremendous improvement between 1950 and 2020 that holds for all countries and gender population and can be seen by the relatively abrupt transition of the color gradient, from red to green. The decline from the color gradient for the year 2020 is especially remarkable for the United States and Italy, probably linked to the comparatively higher COVID-19 incidence and mortality rates in both countries. More specifically, this can be observed by the fact that the color profile on the far right side, has moved vertically downwards, indicating that a relatively sharp decline can be tracked for the year 2020. The area in the lower right corner of the heatmaps indicates a very low mortality rate for ages 1 to 10 in the years after 1950. This might be caused by the improved hygiene and new medical innovations.

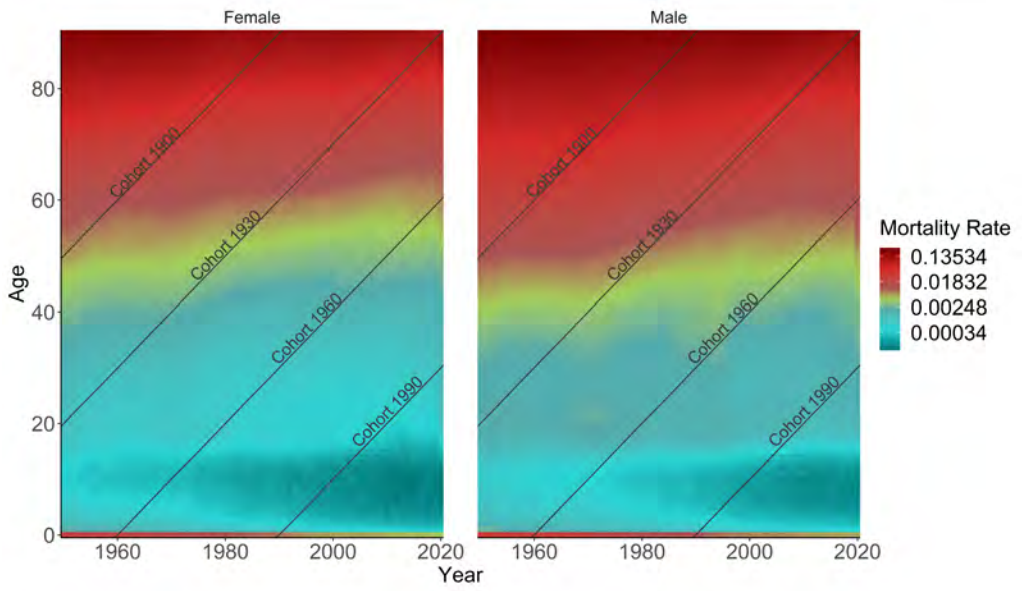


Figure 4: Heatmaps of mortality rates for the US population. Age groups and periods are depicted in vertical and horizontal direction, respectively. Unique cohorts are displayed along the diagonals. It should be noted that the mortality data for 2020 were enriched from the STMF dataset, with the methodology described in detail in Section 3.

Additionally the diagonal lines represent the corresponding birth cohorts: 1900, 1930, 1960 and 1990. Accordingly, unique cohorts are displayed along the diagonals. For instance, cohort 1900 refers to the subset of people born in 1900. A 40−years old US American woman born in 1930 (green) tends to have a higher mortality rate than the other one born in1960 (blue). Furthermore, a 10−years old boy has a much lower mortality rate if he was born in 1990 (dark blue) than the other one born 1960 (light blue). The view from this perspective is generally interesting because cohorts also play a crucial role in the later course, partially in the modelling phase.

3 Data Enrichment

This section is about enriching the existing data described in the upper section. This happens in two dimensions. First, we explain the methodology we used to adapt the mortality data from STMF (2021), namely the number of deaths, but also the population for recent years that are missing in HMD (2021) dataset. The challenge we specifically address is that these data are aggregated in rough age categories, whereas we are interested in the metric age scale. The second dimension of data enrichment refers to the integration of socio-economic and health variables obtained from other public datasets and should be put in relation with mortality data.

3.1 Distribution of population and deaths from categorized into metric age scale

Forrecent years only partial information about mortality rates and population is available on weekly basis and with grouped instead of individual age information (STMF, 2021). This refers to the year 2020 for all countries, year 2019 for Germany and Italy and 2018 for Germany only. Thus, both death counts and population have to be transformed from weekly to yearly data first, and then from age buckets into the metric age scale.

As the STFM-dataset provides weekly information only about mortality rates and death counts, one has to firstly derive the weekly population size from this and then extrapolate to the annual level. From the documentation of this data set, the original calculation of the weekly age-specific mortality rate can be obtained (STMF, 2021):

$$\mu_{[l,u],w,t,s} = \frac{D_{[l,u],w,t,s}}{E_{[l,u],t,s}/52} \tag{2}$$

where $E_{[l,u],t,s}$ represents the annual population exposures, $D_{[l,u],t,s}$ the observed death counts for a given age interval $[l, u]$, in year t, week w for subpopulation s. The population for week w is therefore given as follows:

$$E_{[l,u],w,t,s} = \frac{D_{[l,u],w,t,s}}{\mu_{[l,u],w,t,s}} \tag{3}$$

with $\mu_{[l,u],w,t,s}$ indicating the mortality rates for a given age interval $[l, u]$, in year t, week w for subpopulation s. The weekly population is mapped to the yearly population by multiplying with a factor 52, as an indicative number of weeks in a year:

$$E_{[l,u],t,s} = E_{[l,u],w,t,s} \cdot 52 \tag{4}$$

On the other side, the weekly death counts will be summarized per year:

$$D_{[l,u],t,s} = \sum_{w=1}^{52} D_{[l,u],w,t,s} \tag{5}$$

To construct the annual death counts and populations for individual ages for 2020 and 2018-2019 (if necessary), the procedures described in the next two subsections were applied. While the mapping of population takes advantage of the similarity of the cohort-wise population, the other method used assumes that the distribution of death counts within the age groups remains roughly similar on average to that of the last five years. The methods will be explained with the help of US male population and have been applied to all subpopulations in the same way.

Methodology to construct the population data for each age, for year 2020 (metric level):

The idea is to use the similarity of the cohort-wise patterns of the past years and assume that this is also true for the year 2020, in accordance with the recommendation of Antonio et al. (2020). As shown in figure 5, the population pattern for years 2015-2019 is nearly the same, but shifted to the right by one year. That is natural, because 10−years old boys 2018 belong to the group of 11−years old boys 2019 for example. So, the population of 2019 will be right shifted to create the initial population course for 2020. This leads to an empty observation for 0 age in 2020 (in a first step), which will be linearly extrapolated based on the 0 age population of the last two years (2018 and 2019). Figure 6 shows how the population size of men and women in the United States is distributed. Obviously more people achieved an age above 60 in 2010 than in 1900 because of the triangular blue area in the top left corner. The years of the baby-boom generation starting in 1962 where more babies were born than in any other year can be seen. This is probably the reason why there is a growing number of people between ages 30 and 50 in the years after 1990. However, the boom ended rapidly around 1970 when the birth control pill was invented. This changed the age structure of society from a population pyramid in 1912, where the population size decreased for an increasing age, to a more urn shaped form in 2010, which means that there are fewer people born than having an age from 30 to 50 (Yang and Land, 2013).

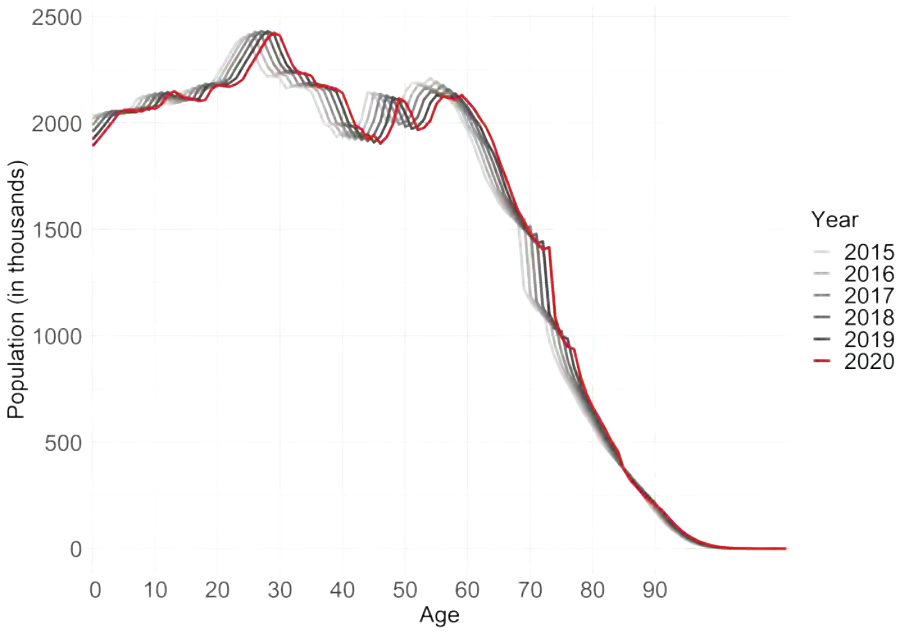


Figure 5: Enriched exposure (red) for US male

The next step aims to ensure that the newly created population numbers (red line in figure 5) matches the observed population numbers within each age group prvodied by STMF (2021) if summed up across the corresponding categories given in the table below:

Age group	Observed	Derived
$[0, 14]$	30.694.119	30.623.608
$[15, 64]$	106.359.573	106.664.997
$[65, 74]$	15.203.936	15.511.873
$[75, 84]$	7.227.623	7.600.156
$[85+]$	2.416.311	2.832.233

Table 3: Exposure of the US male population for specific age groups for the year 2020. Observed values are provided by STMF (2021) and derived values are based on previous year information from the HMD (2021) dataset

To set the two columns to the same common mass, so that the derived values in the sum correspond to the actual values, the following connection will be assumed:

$$\begin{aligned} E_{\{l,u\},2020,s} &= r_{\{l,u\},s} \cdot E^*_{\{l,u\},2020,s}, \\ \text{with } \; r_{\{l,u\},s} &= \frac{E_{\{l,u\},2020,s}}{\sum_{i=l}^{u} E^*_{i,2020,s}} \end{aligned} \tag{6}$$

Here, $E^*_{i,2020,s}$ indicates the exposure size after the shift to right for a given individual age i and subpopulation s for year 2020. With the help of the ratio $r_{\{l,u\},s}$ the sum of initially shifted population counts in the given age groups will be equal to those given in the STMF (2021) dataset for year 2020. Recall, we fixed at the beginning that the maximum age limit being monitored in this thesis is 90. Therefore,

the population in the open interval 85+ must be adjusted, so that the increase from 2019 to 2020 is evenly distributed across 85 to 110:

$$\begin{aligned} E_{\{85,90\},2020,s} &= e_{\{85+,s\}} + E_{\{85,90\},2019,s}, \\ \text{with } \; e_{\{85+,s\}} &= \frac{E_{\{85+\},2020,s} - E_{\{85+\},2019,s}}{110 - 85 + 1} \end{aligned} \tag{7}$$

This allows a closed-interval statement on population trends to be made reliably up to the age of 90. The final result of population values for the year 2020 at a precise age level is depicted in figure 5.

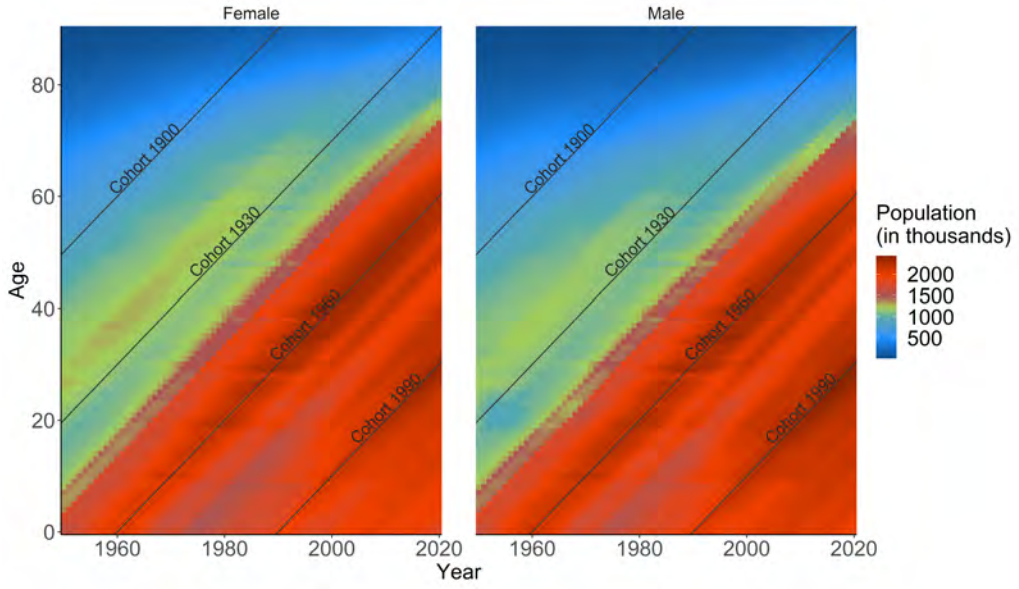


Figure 6: Heatmaps of population numbers for the US population. Age groups and periods are depicted in vertical and horizontal direction, respectively. Unique cohorts are displayed along the diagonals. It should be noted that the mortality data for 2020 were enriched from the STMF dataset, with the methodology described in detail in this section.

Methodology to construct the death counts for individual age for year 2020:

Now, a three-stage approach will be taken to distribute the death counts from grouped into metric age scale on annual level. The core concept is to exploit the assumption of homogenized distribution in the given age buckets over the years 2015-2019 and to map this in an averaged form to the year 2020. Figure 7 illustrates the given issue exemplary for the year 2019: for this year (and also for 2015-2018) both the death counts for individual (colored bars in figure 7) and grouped (on average, black lines in figure 7) ages are available. Thus, for the year 2020 only the information comprised in and corresponding to the black lines are present.

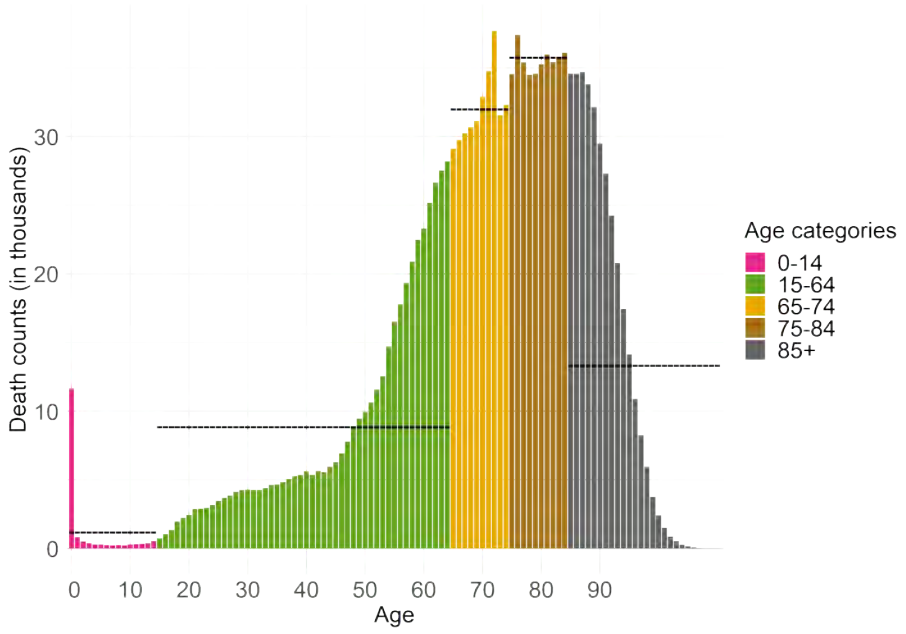


Figure 7: Examplarily depiction of averaged grouped death counts for the US male population. The black lines are obtained from the mortality information in the STMF dataset (STMF, 2021).

Firstly, averaged weights for each age in each age bucket are computed based on the recent five years (2015-2019):

$$w_{\{l,u\},t,s} = \frac{D_{\{l,u\},t,s}}{\frac{1}{u-l+1} \cdot D_{[l,u],t,s}} \tag{8}$$

Secondly, these weights are multiplicatively applied on the averaged death counts 2020 (black lines) in each age group to correct for the deviation from the mean. The assumption made in this step is, that the distribution within each age category remains the same across the previous years. $D^*_{\{l,u\},t,s}$ denotes the estimates death counts for each individual age contained in an age group, year t and subpopulation s:

$$D^*_{\{l,u\},t,s} = (\frac{1}{5} \sum_{j=2015}^{2019} w_{\{l,u\},j,s}) \cdot \frac{1}{u-l+1} \cdot D_{[l,u],2020,s} \tag{9}$$

Finally, the same principle as for the population method is followed to ensure equal death counts in both grouped and metric versions, when summing up within each age group:

$$\begin{aligned} D_{\{l,u\},2020,s} &= k_{\{l,u,s\}} \cdot D^*_{\{l,u\},t,s}, \\ \text{with} \quad k_{\{l,u,s\}} &= \frac{D_{[l,u],2020,s}}{\sum_{i=l}^{u} D^*_{i,2020,s}} \end{aligned} \tag{10}$$

In this way it must be guaranteed that the newly created death counts in total each result in the same number as the grouped data from STMF (2021) given in the following table:

Age group	Observed	Derived
$[0, 14]$	16.219	16.131
$[15, 64]$	543.294	543.829
$[65, 74]$	395.856	396.675
$[75, 84]$	432.905	433.527
$[85+]$	399.138	398.762

Table 4: US male death counts in both datasets for given age groups

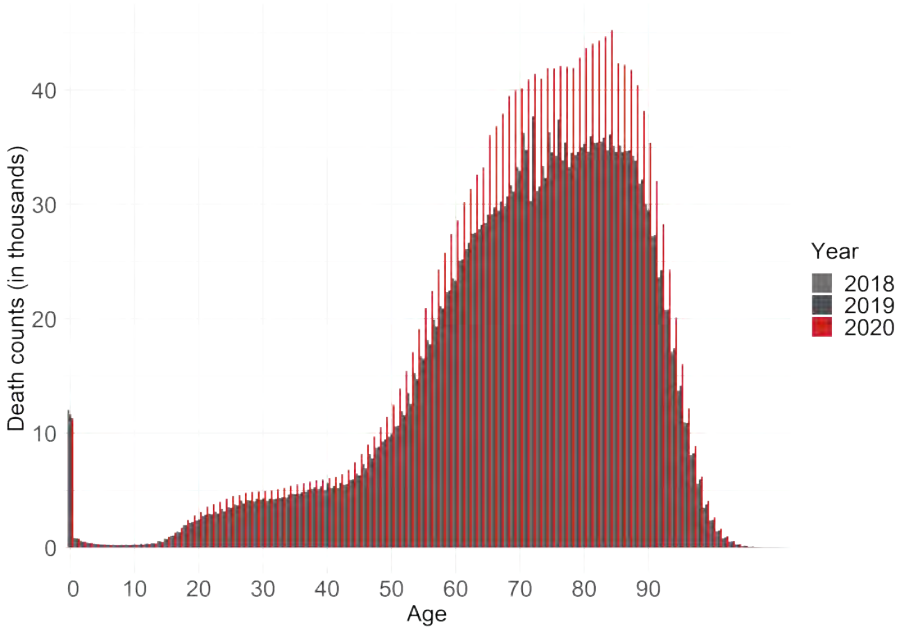


Figure 8: Adjusted death counts (red) for US male

The final death counts for individual ages for years 2018, 2019 (greyish colors) and 2020 (red) are shown in figure 8. For higher age categories the increase in death counts (higher bars) is immediately apparent. Presumably, this is related to the COVID-19 situation in the USA.

Mortality rates are computed as usual by dividing the death count by the population number for each individual age and subpopulation. Finally, the HMD dataset has been imputed by adjustedmortality rates for 2020. An equal procedure was followed for each subpopulation (country and gender) and for all missing years as just described.

3.2 Integration of indicators on socio-economic and health status

Up to this evaluation phase, the data set consists of the following variables: year of death, age (metric)at death, death count, population size, mortality rate gender and country. For the coherent modeling over countries, it might be not sufficient to examine how mortality rates differ across subpopulations in each country, defined by gender and age categories only. Since it seems obvious that the development of mortality rates varies also across different socio-economic circumstances within a country, variables measured by educational attainment, occupation, income or area deprivation could be beneficial to better capture and project the mortality rates. These socio-economic differences in mortality pose a major challenge for the design of pension systems and the management of longevity risk in pension funds and pension portfolios to address social inequalities. It should not be ignored, however, that failure to account for mortality heterogeneity in the valuation of pension obligations or the pricing of annuities could lead to inadequate funding of pension and retirement obligations, thus inducing undesirable transfer of wealth (Brown et al., 2002).

Therefore, it is of interest to include more variables into the existing dataset indicating the socio-economic but also health and educational situation. Current research offers few modern techniques to include socio-economic or even other additional variables. However, these are usually based on individual data and individual-specific information on income, socio-economic and health status (Villegas, 2015). In this thesis, however, we work with aggregated data, refelcting rather the differences between countries.

The described enrichment will be achieved by extracting this information from additional sources, such as the Organisation for Economic Co-operation and Development (OECD, 2021), World Health Organisation (WHO, 2021), United Nations Development Programme (UNDP, 2021), the World Bank (World Bank, 2021), UNESCO Institute of Statistics (UIS, 2021). As such public data sources usually do not provide any gender- and age-specific information, the inclusion of these variables into the dataset and into the models can be considered as a sort of explorative analysis. In general these are rather an indicator for the lifestyle in the country in a specific year and should therefore not be viewed as direct influencing factors on the mortality rate. The usefulness and beneficialness of this kind of reflection of differential mortalities and comparative performance of countries as well as major trends will be assessed in detail in Section 5.

From a total of about 1500 indicators collected from different domains and sources, four were selected. The variable and domain selection were mainly based on correlation check, expert suggestions and content assessment.

1. Alcohol consumption measured in litres per person for adults, namely people aged 15 years and older. It is defined as yearly sales of pure alcohol in litres per person. In general, alcohol consumption is linked to several harmful health and social consequences, particularly an increased risk of a number of cancers, stroke, and cirrhosis of the liver (OECD, 2020a).

2. Prevalence of overweight among adults measured by the Body-Mass-Index (BMI) as a risk factor for health. This gives the percentage of people aged 15 or

higher with a BMI of $25kg/m^2$ or higher. BMI is commonly used to categorize overweight and obesity and is defined as a person's weight in kilograms divided by the square of his height in meters (WHO, 2020).

3. Tax revenue measured as tax to gross domestic product (GDP) ratio determining how well a nation's government directs its economic resources via taxation. The value is given as percentage. This should serve as a performance indicator for a prosperous national economy. Higher tax revenues mean that a country is able to invest more in improving infrastructure, healthcare and education. This indicates key conditions to the long-term prospects of a country's economy and people (OECD, 2020b).

4. Unemployment rate is measured by the number of unemployed people as a percentage of the labour force and it is seasonally adjusted. Unemployed is hereby defined as people of working age without work, available for work and have taken specific steps to find work. Data are based on labour force surveys (OECD, 2020c).

Alcohol consumption (in litres per person)								
Country	Minimum	1^{st} Quartile	Median	3^{rd} Quartile	Maximum	Mean	Std. deviation	Missing
Finland	8,2	8,5	9	9,7	10,5	9,0966	0,6941	2
Germany	10,8	11,2	11,8	13	14,9	12,1759	1,1115	2
Italy	7	7,4	8,9	9,6	11	8,6862	1,2463	2
Netherlands	8,2	9,1	9,7	9,9	10,1	9,4345	0,6269	2
United States	8,1	8,3	8,6	8,725	9,3	8,5536	0,2987	3
Prevalence of overweight (in %)								
Country	Minimum	1^{st} Quartile	Median	3^{rd} Quartile	Maximum	Mean	Std. deviation	Missing
Finland	48,3	53,35	57,1	60	62,5	56,4963	4,2464	4
Germany	48,5	52,9	56,6	59,7	62,8	56,2333	4,2852	4
Italy	48,8	52,65	57	60,75	64,1	56,7111	4,7914	4
Netherlands	43,2	48,75	54,2	58,95	62,6	53,7037	6,0981	4
United States	51,4	57,2	62,4	66,6	70,2	61,7592	5,7619	4
Tax revenue (in %)								
Country	Minimum	1^{st} Quartile	Median	3^{rd} Quartile	Maximum	Mean	Std. deviation	Missing
Finland	40,562	42,111	43,195	44,1965	45,759	43,1555	1,4355	1
Germany	34,329	35,1828	35,8125	36,7795	38,812	36,0414	1,1769	1
Italy	36,271	39,6943	40,658	41,9018	43,828	40,7157	1,8656	1
Netherlands	34,804	35,623	36,8525	38,627	41,728	37,1932	2,0766	1
United States	22,957	24,8253	25,9335	26,7618	28,291	25,843	1,3547	1
Unemployment rate (in %)								
Country	Minimum	1^{st} Quartile	Median	3^{rd} Quartile	Maximum	Mean	Std. deviation	Missing
Finland	3,1667	7,7667	8,7667	10	16,5833	9,4366	3,0116	0
Germany	3,15	5,2771	7,7167	8,6688	11,2833	7,1542	2,3157	1
Italy	6,15	8,4917	9,725	11,175	12,825	9,6753	1,7461	0
Netherlands	3,075	4,2042	5,1333	6,275	8,3583	5,3685	1,4040	0
United States	3,667	4,675	5,6083	6,8583	9,6167	5,914	1,6198	0

Table 5: Summary statistics on the values of relevant indicators

Table 5 gives an overview of the index values for each country. The data basis for this table is the annual data for the four indicators per country. The time frame covers the years 1990-2020. In addition to the minimum, maximum and mean values per country and indicator, the first and third quartiles, the standard deviation and the number of missing values are also shown. The prevalence of overweight among adults increases over the years in all countries. The USA stands out with the maximum value of BMI and the minimum value of tax revenue. However, the data is available for the period range of 1900-2020 and contains missing values, which have

to be imputed with the help of `imputeTS` package (Moritz, 2021). This provides an imputation option for univariate time series, which is required in this case to capture the time trend in the data and thus project the time series nature. The application of Kalman Smoothing to the state-space representation of an Autoregressive integrated moving average (ARIMA) model for each country and each indicator is shown in figure 9. For the exact procedure in this matter, reference is made to the following sources (e.g. Welch and Bishop, 2006, Harvey, 1990, Grewa, 2011). The results appear to be satisfactory, as they plausibly extend the trend present in the data. However, due to the pandemic influencing the year 2020 in particular in its special way, the reproducibility of the previous trends is questionable, as the development of these indicator values would presumably have been different from what it would have been if there had been no pandemic. Therefore, these covariates are only integrated into the modeling up to the year 2015 and 2019, respectively, so that it is not a further problem in the course of this thesis.

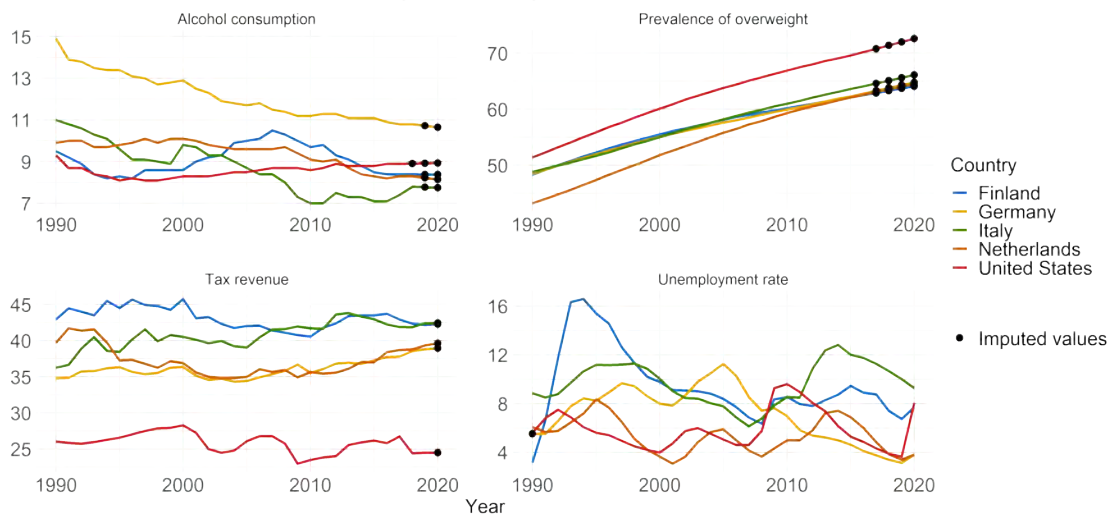


Figure 9: Socio-economic and health-related covariates as time series with imputed values

While the objective of Kalman Smoothing is to estimate the unknown state vector of the structural model given the entire sample, the automate version of ARIMA modelling returns the optimal model, according to AIC. It helps to cope with the order selection process, usually considered to be subjective and difficult to apply, will be discussed in detail in Section 4 (Hyndman and Khandakar, 2008, Kalman, 1960). Another imputation approach called `MissForest` was also taken (Stekhoven, 2016). However, this method builds upon Random Forest Algorithm and does not consider the underlying time-dependent structure leading to non-satisfactory results, especially for variable with a trend, which will be ignored.

All in all, the above methods allow for extensive enrichment of the initial data, both in terms of adding annual age-specific mortality data for the missing years and incorporating more covariates from additional data sources. And these covariates are used for later modelling in Section 5 where their relationships with mortality are examined in more detail.

4 Methods

Both the level of individual and collective well-being in societies is a matter of public, political and scientific interest, intensified due to the COVID-19 pandemic and its hardly predictable developments. As observed in the past, life expectancy in a society generally increases in a monotonic pattern with disruptive spikes. This justifies the approach of modelling human mortality rates with stochastic models assuming stochastic processes behind the evolution over time and not just considering fixed rates. In most cases, we take into account the minimal characteristics with age, cohort and gender effects. This serves as a basis for considering different models with strengths and weaknesses. In this section the covered models are presented. The focus is firstly on the comparison of different models with the help of appropriate metrics as well as in-sample and out-of-sample forecasts. First, an overview of Lee Carter models and its variants will be given. Apart from the model specification and general idea, the methodology behind the modelling of structural breaks will be explained. Then, standard machine learning algorithms will be presented and then linked to the Lee Carter model as an improvement strategy. Lastly, generalized additive model will be introduced and put into the context of modelling and extending the well-known Age-Period-Cohort (APC) models. A uniform notation was chosen, which in individual cases differs from original notation in the sources.

4.1 Lee Carter and variants

One of the most comprehensible approaches for modelling mortality in a set of subpopulations is the classical version of Lee Carter model (Lee and Carter, 1992), further adapted by the Poisson distributional assumption (Brouhns et al., 2002). The original version of Lee Carter model proposed by Lee and Carter (1992) suggests to forcast the mortality rates $\mu_{a,t}$ at age a in year t based the following formulation:

$$\log \mu_{a,t} = \alpha_a + \beta_a \kappa_t + \epsilon_{a,t} \tag{11}$$

α_a represents the average shape of the overall age course, β_a describes the improvement rate of mortality at age a and κ_t the general trend of mortality at time t. The error terms $\epsilon_{a,t}$ reflect the residual age-specific historical influence on mortality rates that the model is not able to capture. The authors propose to use the singular value decomposition (SVD) method to find the least squares solution to the minimization problem of the residual sum of squares considering the described in the below sections that are imposed to counteract the identifiability issue:

$$\sum_{a,t} (\log \mu_{a,t} - \alpha_a - \beta_a \kappa_t)^2 \tag{12}$$

With this perspective, the approach can be characterized as follows: the estimated $\hat{\alpha}_a$ is equal to the average of the observed logarithmic mortality rates $\log \mu_{a,t}$ over time of an age profile a, yielding in $\overline{\log \mu_a} = \hat{\alpha}_a$. The first term of the SVD of the matrix $\log \mu_{a,t} - \hat{\alpha}_a$ is then described by the $\hat{\beta}_a$ and $\hat{\kappa}_t$. In order to make the resulting fitted deaths to match the total number of the actual deaths observed in each year t,

a second-stange estimation is adopted to adjust the parameter κ_t. With this variant of the model-setup the errors assumed to be homoskedastic and normally distributed which is required for being able to perform the least squares estimation via SVD. The fixed effects model for the centered age profile can be therefore reformulated as $\log \mu^*_{a,t} := \log \mu_{a,t} - \hat{\alpha}_a = \log \mu_{a,t} - \overline{\log \mu_a}$ and thus with the normality assumption resulting in the model formulation:

$$\log \mu^*_{a,t} \sim \mathcal{N}(\overline{\log \mu_{a,t}}, \sigma^2) \tag{13}$$

$$\mathbb{E}(\log \mu^*_{a,t}) = \beta_a \kappa_t \tag{14}$$

The matrix elements below illustrate that for this expression the feature space to be estimated comprises only $A \times T$:

$$\log \mu_* = \begin{array}{c} \\ \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{90} \end{array} \begin{array}{c} \begin{array}{ccccc} \kappa_{1950} & \kappa_{1951} & \kappa_{1952} & \cdots & \kappa_{2020} \end{array} \\ \begin{pmatrix} \log \mu^*_{0,1950} & \log \mu^*_{0,1951} & \log \mu^*_{0,1952} & \cdots & \log \mu^*_{0,2020} \\ \log \mu^*_{1,1950} & \log \mu^*_{1,1951} & \log \mu^*_{1,1952} & \cdots & \log \mu^*_{1,2020} \\ \log \mu^*_{2,1950} & \log \mu^*_{2,1951} & \log \mu^*_{2,1952} & \cdots & \log \mu^*_{2,2020} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \log \mu^*_{90,1950} & \log \mu^*_{90,1951} & \log \mu^*_{90,1952} & \cdots & \log \mu^*_{90,2020} \end{pmatrix} \end{array}$$

The expected cell value can be calculated by the product of the two parameters from the correspoding marginals. For instance, for age 2 and year 1952, the Lee Carter approximates by the following expression $\log \mu^*_{2,1952} = \beta_2 \kappa_{1952}$. Thus, for a given age a, this model can be viewed as similar to a log-linear model with the period effect κ_t serving as a covariate and the mortality improvement rate β_a being equivalent to the slope (Girosi and King, 2007). The deviation from the classical log-linear model is the fact that these are unobserved variables, which is why the problem here is not a vector estimation but a matrix estimation. A one rank approximation tries to estimate this matrix, hence SVD is used. As the logarithm of the observed mortality is much more variable at older than at younger ages due to the much smaller death counts, the assumption of homoskedastic and normally distributed errors is considered unrealistic (Brouhns et al., 2002). In order to address this limitation Brouhns et al. (2002) suggest to put the LC model into a setting with Poisson distributional assumption, presuming that the observed death counts $D_{a,t}$ follow a Poisson distribution at a given age a, in year t. This serves as a benchmark model, allowing further development to be carried out and compared on the grounds of this model. This idea is exploited by the generalized nonlinear models and applied in the background implementation of the package `StMoMo` used in this thesis (Villegas et al., 2016).

Even though the notation for subpopulation s has been omitted up to this point for simplicity, LC and the variants are basically individual modeling of subpopulations. In this case of mortality modelling each individual person is identified by a feature tuple $s \in S = (gender \times country)$, with $gender\ G = \{female, male\}$ and $country\ C = \{FIN, DE, ITA, NLD, US\}$ describing the subpopulation as well as

age $A = \{0, \ldots, 90\}$ and period which indicates time in years $T = \{t_{min}, \ldots, t_{max}\}$. The exposure $E_{a,t,s}$ and the corresponding number of deaths $D_{a,t,s}$ will be assigned to a given age a, year t and subpopulation s.

Model Specification:

The Lee Carter model under the setting of Poisson distributional assumption is specified as follows:

$$D_{a,t,s} \sim Poisson(E_{a,t,s} \cdot \mu_{a,t,s}), \text{ independent distributed} \tag{15}$$

$$\eta_{a,t,s} = \alpha_{a,s} + \beta_{a,s}\kappa_{t,s} \tag{16}$$

The link between the linear predictor $\eta_{a,t,s}$ and the mortality rates can be described like this: $\log \mu_{a,t,s} = \eta_{a,t,s}$. Hereby again, $\alpha_{a,s}$ captures the general age-specific mortality pattern for subpopulation s, $\kappa_{t,s}$ is the time varying mortality index representing the overall level of mortality in year t for a given subpopulation s and $\beta_{a,s}$ measures the age-specific response to changes in the general level of $\kappa_{t,s}$.

Identifiability constraints:

As introduced in the original paper (Lee and Carter, 1992) the model contains identifiability problems. More precisely, it means that the model parameters are only identifiable up to a transformation, so that values of different effects are not identified in a unique manner. Choosing c_1 and c_2 not equal to zero arbitrarily as real constants, the parameters from the model specification can be transformed in this way leading to the same mortality rates: $(\alpha_{a,s}, \quad \beta_{a,s}, \quad \kappa_{t,s}) \rightarrow (\alpha_{a,s} + c_1\beta_{a,s}, \quad \frac{1}{c_2}\beta_{a,s}, \quad c_2(\kappa_{t,s} - c_1))$. Therefore, the following constraints are incurred, in order to guarantee the identifiability of the model: $\sum_{a \in A} \beta_{a,s} = 1$, $\sum_{t \in T} \kappa_{t,s} = 0$. The function that defines the identifiability constraints of the model is implemented in the StMoMo package to take a set of fitted model parameters and return a list of the transformed model parameters with the identifiability constraints function applied. For instance, in case of Lee Carter model, the transformation of the above vector representation can be imposed by choosing the following constraints and using each estimated parameters: $c_1 = \frac{1}{n}\sum_t \kappa_t$ and $c_2 = \sum_a \beta_a$.

Since the time series forecasting method ARIMA will be mentioned frequently in the further course of this thesis, the term with its theoretical background will be introduced in a short form at this point. One way of forecasting trend, age, country and gender specific mortality rates is to use the classic ARIMA model (Box and Jenkins, 1976), which is sufficient and appropriate if the available database is large enough and provide a stable and consistent pattern over time with only few outliers. The common notation of an ARIMA(p,d,q) model includes the order p of the autoregressive process, the order d of integration, namely the number of times that the series must be differenced in order to make it stationary and the order of the moving average process q. Its detailed form for a stochastic process $\mu_{a,t,s}$ is set up in the following but will only be presented with reduced notation in the further course:

$$\Delta^d \mu_{a,t,s} = \overbrace{\delta_{a,s}}^{\text{Drift}} + \overbrace{\sum_{i=1}^{p} \phi_i \Delta^d \mu_{a,t-i,s}}^{\text{Regression}} + \overbrace{\epsilon_{a,t,s} + \sum_{j=1}^{q} \theta_j \epsilon_{a,t-j,s}}^{\text{Smoothed noise}} \tag{17}$$

The idea hereby is to obtain the response from the linear regression of previous differences and additional smoothed noise. We denote with $\Delta^d \mu_{a,t,s}$ the stationary time series. These are then used to fit the ARIMA model. The drift $\delta_{a,s}$ is a constant parameter and indicates the average change in the series over time. Moreover, ϕ_i are the parameters of the auto-regressive part and the θ_j the parameters of the moving average part. The sequence of independently distributed random variables is normally distributed with $\epsilon_{a,t,s} \sim \mathcal{N}(0, \sigma)$ (Pascariu, 2018).

Parameter Estimation:

To fit the parameters $\alpha_{a,s}$, $\beta_{a,s}$, $\kappa_{t,s}$ to the data the R package `StMoMo` was used (Villegas et al., 2016). This implementation makes use of the so-called Generalised Age-Period-Cohort (GAPC) stochastic mortality modelling. According to Brouhns et al. (2002) the number of deaths, say the random component, follows a Poisson distribution with a parameter proportional to the exposure as follows: $\mathbb{E}(D_{a,t,s}/E_{a,t,s}) = \mu_{a,t,s}$, where the $\mu_{a,t,s}$ denotes the underlying mortality rate specified in the equations above. Moreover, $D_{a,t,s}$ are independent in $(a, t, s) \in (A \times T \times S)$. With this assumption, a standard Maximum Likelihood approach is used for the estimation of parameters after choosing the log link function for the mortality rates $\mu_{a,t,s}$. Let $d_{a,t,s}$ be the corresponding number of deaths actually observed. so that the log Likelihood function can be formulated down as follows:

$$log(\mathcal{L}(\alpha_{a,s}, \beta_{a,s}, \kappa_{t,s})) = \sum_{a,t,s} d_{a,t,s}\, log(E_{a,t,s} \cdot \mu_{a,t,s}) - E_{a,t,s} \cdot \mu_{a,t,s} - log(d_{a,t,s}\,!) \tag{18}$$

whereas $D_{a,t,s}$ will be seen as a counting random variable with $D_{a,t,s} \sim Poisson(E_{a,t,s} \cdot \mu_{a,t,s})$ and $\mu_{a,t,s} = e^{\alpha_{a,s} + \beta_{a,s}\, \kappa_{t,s}}$. In the literature this perspective of viewing is also called Lee Carter model under a Poisson setting (Villegas, 2015). Following the structure specifically for the Lee Carter model suggested originally by Lee and Carter (1992), the predictor or the so-called systematic component results in $\eta_{a,t,s} = \alpha_{a,s} + \beta_{a,s}\kappa_{t,s}$. The link function g connects the random component and the systematic component: $g(\mathbb{E}(D_{a,t,s}/E_{a,t,s})) = \eta_{a,t,s}$. Choosing the log function as link, the expression yields in $\log \mu_{a,t,s} = \eta_{a,t,s}$.

To forecast mortality, the period index κ_t (the index s for the subpopulation is omitted for reasons of clarity) will be modelled and forecasted into the future using univariate ARIMA processes. This will be implemented with a random walk with drift (ARIMA(0,1,0)) as suggested by the authors (Lee and Carter, 1992):

$$\kappa_t = \delta + \kappa_{t-1} + \xi_t \tag{19}$$

with $\xi_t \sim \mathcal{N}(0, \sigma_\kappa^2)$ i.i.d.. Hereby, δ denotes the drift term, ξ_t the error term which is normally distributed with zero mean and variance σ_κ^2. It is worth mentioning that the error terms are assumed to be independent, as the modelling is done independently for each subpopulation. The validity of this assumption is however questionable. The reason for this is that primarily subpopulations in relation to gender within one country tend to have similar mortality trends. In addition, it is also possible that similarities and dependencies in mortality exist between countries, for example from a geographical or political angle. Another assumption made implicitly is that the mortality rates $\mu_{a,t,s}$ stay constant over the period $(t, t+1]$.

Structural breaks:

The presence of structural changes can prevent the model from achieving satisfactory capturing of the time-dependent effects, in this case period effect κ_t, due to the sensitivity of ARIMA modelling. This might of course also affect the mortality forecasting results which are very delicate towards the calibration period. To investigate the structural breaks in the period effect, we first applied the Lee Carter model. Then extracted the estimated values for κ_t and calculated the first degree differences. This provides the time series as the basis for detecting the structural breaks. Modelling structural changes for period effect can be introduced in two dimensions: Apart from dating the structural changes, the optimal number of break points needs to be determined. The aim is therefore to find the best partition (break dates) of a time series for an a priori given optimal number of break points k^*. This means that the underlying range of year values will be partitioned into $k^* + 1$ segments. Then, for each partition, the least squares estimates of the regression coefficients are obtained within the framework of a piecewise linear regression model. Hereby, the first order differences of the period effects $\kappa_{a,t,s}$ are considered as dependent variable. Based on this, the minimal residual sum of squares results in the following estimated breakpoints:

$$(\hat{p_1}, \ldots, \hat{p_k}) = \underset{p_1,\ldots,p_k}{\operatorname{argmin}} RSS(p_1, \ldots, p_k) \tag{20}$$

$$RSS(p_1, \ldots, p_k) = \sum_{j=1}^{k+1} rss(p_{j-1} + 1, p_j) \tag{21}$$

Here, $RSS(p_1, \ldots, p_k)$ defines the minimal residual sum of squares in the j^{th} segment for the breakpoints $(p_1, \ldots, p_k)$, whereas p_0 and p_{k+} indicate the beginning and the end of the given time series, respectively. The main idea is to minimize the following objective function for every possible partition $(p_1, \ldots, p_k)$. Yet, obtaining the global minimizer requires high, computationally intensive capacities, because the number of possibles segments (and thus also computation times of sum of squared residuals) reaches $n(n+1)/2$ for a sample size of n. This problem is overcome by applying a dynamic programing algorithm based on a recursive procedure. This technique basically proceeds by sequentially inspecting optimal one-break partitions. It exploits the idea of Bellman's principle which states that it is enough to know the "optimal previous partner" for each point i if i was the last breakpoint in a k^*-partition. It simply minimises the sum of squared residuals over the set of options

obtained by supposing that the first elements were optimally partitioned into k^* parts and that $k^* + 1$ part consists of the remaining elements of the last procedure. The implementation and methodology details are documented by Zeileis (2005) and Bai and Perron (2003).

One of the basic possibilities is about to choose the optimal number of breaks based on an information criterion given a set of possible values. For this thesis Bayesian Information Criterion (BIC) was used, defined by $BIC = \nu log(n) - 2log(\hat{\mathcal{L}})$, where ν is the number of parameters estimated by the model, n the number of observations and $\hat{\mathcal{L}}$ is the maximized value of the likelihood function of the model. This choice is justified according to Bai and Perron (2003) stating that BIC is suitable in many cases whereas Akaike information criterion (AIC) tends to overestimate the number of breakpoints. Usually such information criteria are used for model selection. In this case a number out of $k = \{0, ..., 5\}$ will be chosen. The implementation was mainly inspired by Zeileis et al. (2003) and Zeileis (2005). The basic idea is to go for k which induces the minimal BIC: $k^* = argminBIC(k)$. Another approach is to test whether the fit improves significantly or not with the help of F-Tests. In this thesis, BIC selection is the course of action we have decided to take, as recommended by Yao (1988) stating that this is a consistent estimator of the true number of break points. Both parts were implemented using the `strucchange` package in R (Zeileis et al., 2002). After detecting the years where structural changes were found, the time frame for the fitting was shortened by the years prior to it.

Variants:

Further various specifications and extensions of the Lee Carter exist. These all have the Poisson distribution assumption in common for death counts proposed by Brouhns et al. (2002). Three of them will be considered, analysed and evaluated in this thesis. In the following, the theoretical foundations are explained building upon the Lee Carter (LC) model.

Age-Period-Cohort model:

The Age-Period-Cohort (APC) model extends the LC model by including a cohort effect $\gamma_{t-a,s}$ and omitting the age-specific improvement rates. The cohort is generally computed by $cohort = year - age$. The model predictor has the following expression:

$$\eta_{a,t,s} = \beta_{a,s} + \kappa_{t,s} + \gamma_{t-a,s} \tag{22}$$

The application of this model has its origins in the field of medicine and demography and goes back a long way (Clayton and Schifflers, 1987, Hobcraft et al., 1982). However, Currie (2006) was the first who considered this type of model in the actuarial field. With the Poisson distribution assumption and the log link function remaining the same it can be traced back to the general shape of Generalized Age-Period-Cohort models (Villegas, 2015). The identifiability can be ensured with the following constraints: $\sum_t \kappa_{t,s} = 0$, $\sum_{c=t_{min}-90}^{t_{max}-0} \gamma_{c,s} = 0$, $\sum_{c=t_{min}-90}^{t_{max}-0} c\gamma_{c,s} = 0$, indicating that the cohort effect oscillates around zero, with no apparent linear trend.

Cairns-Blake-Dowd model:

Another variant is the proposal of (Cairns et al., 2006), which omits a static age function and a cohort effect. Instead, the so-called Cairns-Blake-Dowd (CBD) model integrates the second period term accounting for differences from the average age $\bar{a}$ in the data. This yields in the following model predictor:

$$\eta_{a,t,s} = \kappa_{t,s}^{(1)} + (a - \bar{a})\kappa_{t,s}^{(2)} \tag{23}$$

Moreover, the numbers (1) and (2) stand as indices to distinguish the two period effects from each other. Since this model is identifiable, no parameter constraints are imposed.

M7 – Quadratic CBD model with cohort effect:

The original CBD model will be extended by a cohort effect and a quadratic age effect as suggested by Cairns et al. (2009). The model predictor results in the following expression:

$$\eta_{a,t,s} = \kappa_{t,s}^{(1)} + (a - \bar{a})\kappa_{t,s}^{(2)} + \left((a - \bar{a})^2 - 1/91 \sum_{i=0}^{90}(i - \bar{a})\right)\kappa_{t,s}^{(3)} + \gamma_{t-a,s} \tag{24}$$

Here, an identifiability problem is present again, which makes it necessary to configure a set of parameter constraints to cope with this issue: $\sum_{c=t_{min}-90}^{t_{max}-0} \gamma_{c,s} = 0$, $\sum_{c=t_{min}-90}^{t_{max}-0} c\gamma_{c,s} = 0$, $\sum_{c=t_{min}-90}^{t_{max}-0} c^2\gamma_{c,s} = 0$. Again, the constraints ensure that the cohort effect oscillates around zero, with no apparent linear or quadratic trend and the top right attached indices are to distinguish the three period effects from each other. For all the models holds (here again for clarity reasons the condition on s is omitted): in order to forecast the mortality the period effects will be projected into the future according to the standard approach with a multivariate random walk with drift assuming that:

$$\boldsymbol{\kappa_t} = \boldsymbol{\delta} + \boldsymbol{\kappa_{t-1}} + \boldsymbol{\xi_t^{\kappa}}, \boldsymbol{\kappa_t} = \begin{pmatrix} \kappa_t^{(1)} \\ \vdots \\ \kappa_t^{(N)} \end{pmatrix}, \boldsymbol{\xi_t^{\kappa}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{25}$$

where N denotes the number of period effects in the model specification and $\boldsymbol{\delta}$ is an N-dimensional vector of drift parameters, $\boldsymbol{\Sigma}$ is the $N \times N$ variance-covariance matrix of the multivariate white noise $\boldsymbol{\xi_t^{\kappa}}$. In case of one period effect, the approach with multivariate random walk with drift will be followed (like in LC model).

To capture the dynamics of the cohort effects, an independence from the period effect will be assumed following previous research work by (Cairns et al., 2006, Haberman and Renshaw, 2011, Lovasz, 2011). This allows modelling the cohort effect with a univariate ARIMA(p, q, d) process with drift. Specifically, for the two

models presented here that include a cohort effect, we follow the specific suggestions for parameter combinations presented by Villegas (2015) ARIMA(1, 1, 0) with drift for the APC model and ARIMA(2, 0, 0) with non-zero intercept for the M7 model. For implementation purposes the package `forecast` was used to forecast mortality rates (Hyndman and Khandakar, 2008).

As we will also see later in the results section, the CBD and M7 models have been particularly designed to fit higher ages because they make use of the linearity of the age effects at pensioner ages, which does not hold for lower age groups (Villegas and Haberman, 2014). However, the LC as well as APC models usually provide a good fit over a wide age range and are able to pick up even small non-linearities in mortality curve. This will be confirmed in the further course of the analysis, justifying the focus on these two models hereinafter with the aim of further improvement and extension (van Berkum et al., 2013).

4.2 Machine learning approach

Current research shows that applications of machine learning in the study of the changing structure of populations have been rather limited. Levantesi and Pizzorusso (2019) suggest that this lack of interest is due to the fact that machine learning models are often seen as "black boxes" and that many argue that their results are difficult to explain and interpret. Even though the original Lee Carter model has already been extended and further developed in various ways, as seen in the previous sections and literature, there are still weaknesses regarding the fitting of the mortality rates. In this subsection, the benefit of machine learning algorithms will be illustrated to improve the fit of parametric mortality models, exemplarily the one of the Lee Carter model, by detecting the weaknesses of the original model fit. Apart from that, this approach allows for including further variables, such as education, income, or socio-economic status leading to (hopefully) even more enhancement. It should be made clear from the beginning that the idea is not to replace the traditional methods of mortality modeling, but to support and enrich the existing methods with machine learning approaches. As a reminder, Lee Carter is a model for estimating mortality rates with the following assumptions:

$$D_{a,t,s} \sim Poisson(E_{a,t,s} \cdot \mu_{a,t,s}), \text{ independent distributed} \tag{26}$$

$$\begin{aligned} &\eta_{a,t,s} = \alpha_{a,s} + \beta_{a,s}\kappa_{t,s}, \\ &\text{with} \quad \log \mu_{a,t,s} = \eta_{a,t,s} \end{aligned} \tag{27}$$

Thus, fitting a Lee Carter model means basically to provide estimates for mortality rates using the two inputs exposure $E_{a,t,s}$ and death counts $D_{a,t,s}$. These estimates will be marked as $\mu^{LC}_{a,t,s}$ in this subsection. The expected number of deaths according to the Lee Carter fit can be calculated as follows:

$$D^{LC}_{a,t,s} = E_{a,t,s} \cdot \mu^{LC}_{a,t,s} \tag{28}$$

This model will be evaluated by analyzing whether the constant factor $q_{a,t,s}$, which is to be estimated hierarchically in the second step, is equal to 1 or not:

$$\begin{aligned} D_{a,t,s} &\sim Poisson(d_{a,t,s} \cdot q_{a,t,s}), \\ &\text{with } d_{a,t,s} = E_{a,t,s} \cdot \mu^{LC}_{a,t,s} \end{aligned} \tag{29}$$

whereas, if $q_{a,t,s} \equiv 1$ then the mortality rates estimated by the Lee Carter model are an appropriate choice and there is nothing to be improved by machine learning methods. In case of underestimating the mortality rates in the first stage for a given age, year and subpopulation, the factor q should be increased. According to the same logic, q should be decreased, if the Lee Carter model overestimates the mortality rates (Deprez et al., 2017, Levantesi and Pizzorusso, 2019). To calibrate q, the following three algorithms will be applied and compared: Decision Trees, Random Forest and Gradient Boosting Machine. The machine learning framework will be implemented using the ratio between the observed death counts and the corresponding $d_{a,t,s}$ as response. The inputs are age, cohort and period (year). Even though the cohort effect may pose identifiability issues, the authors justify this choice by appropriate predictions. Further implementation details are explained in Subsection 5.3. The machine learning improved mortality rates are then given by $\mu^{ML}_{a,t,s} = \mu^{LC}_{a,t,s} \cdot q^{ML}_{a,t,s}$. To forecast these mortality rates an ARIMA model with automated parameter estimation will be chosen (Hyndman and Khandakar, 2008). In the following the relevant functionality of the applied tree-based machine learning algorithms will be introduced.

Decision Trees:

Decision trees are non-parametric supervised learning methods that can be applied to both regression and classification problems aiming for a model to predict the outcome variable by learning simple decision rules. These rules are implied from the features in the training data. They are based on the successive rectangular partitions of the feature space via a sequence of binary splits and can be seen as a piecewise constant approximation of the underlying true function. During training, observations are forwarded along the resulting tree structure until they end up in a particular leaf node. This set of splitting rules is summarized in a tree (Hastie et al., 2016). Predictions of the response for a given observation with a certain feature space are made by using the average of the training observations in the region that the observation falls into (James et al., 2017).

Let Θ be the number of partitions and $R_1, R_2, \ldots, R_\Theta$ the terminal regions of the feature space. Then, the decision tree estimator yields in the following expression:

$$\hat{q}_{a,t,s} = \sum_{\theta=1}^{\Theta} \bar{q}_{a,t,s} \mathbb{1}_{\{(a,t,s) \in R_\theta\}} \tag{30}$$

where $\bar{q}_{a,t,s}$ is the prediction for a given observation in leaf node $\theta \in \Theta$. The algorithm is designed in a way, that each split is loss-minimal across all child nodes requiring full evaluation of all possible split and threshold configurations. For further details on the Classification and Regression Trees (CART) algorithm, the reader is

referred to (Breiman et al., 1984). To implement decision trees in R the package `rpart` was used (Therneau and Atkinson, 2017).

In general, trees are easy to interpret and have a facile handling of all feature types, including interactions (the latter ability also holds for Random Forest). They are able to fit perfectly every pattern in the training data if allowed for full growth. Overfitting issues are posed, since they can lack robustness and are highly sensitive to data modifications. Thus, probably the most important drawback of Decision Trees is its low-bias and high-variance nature.

Random Forest:

To overcome the overfitting issue, an aggregation of many decision trees was introduced with the hope to correct for the bias-variance trade-off. This however only holds for the case of uncorrelated trees to ensure an improvement of predictive performance of trees. One can imagine this procedure in two parts: First, the so-called bagging, standing for bootstrap aggregation, will be applied in order to bootstrap (sample with replacement) B training samples from the original training data and obtain tree base learners for each that are assembled to predict the response value via averaging. Second, only a random set of features may be considered and used at each split in the tree (Breiman, 2001). This contributes to the decorrelation and yields in relatively stable ensembles, giving the chance to all predictors to be chosen as split candidates, even in the precense of a strong predictor in the training dataset. Proceeding from the notation given for the Decision Trees, the Random Forest estimator is calculated as follows:

$$\hat{q}_{a,t,s} = \frac{1}{B} \sum_{b=1}^{B} \hat{q}_{a,t,s}^{(b)} \tag{31}$$

To implement Random Forest in R the package `randomForest` will be obtained (Liaw and Wiener, 2018). The number of classification trees and thus of bootstrap samples B, as well as the maximum number of levels of the trees, are to be determined in advance. It should be considered that the increase of B does not lead to overfitting problems due to the nature of bootstrapping (James et al., 2017). However, above a certain number, little or no improvement can be expected. Due to the random selection of the m covariates, but also of the individual bootstrap subsets, the Random Forest algorithm usually exhibits a high prediction accuracy and robustness. Moreover, it often provides satisfactory results in the case of high-dimensional data (Strobl et al., 2009). Furthermore, since the bootstrap method uses on average 2/3 of the training data for tree construction, it is possible to use the remaining so-called out-of-bag data for the evaluation of prediction accuracy. A disadvantage of Random Forest is that the computation time is relatively long due to the multiple trees. With the help of an ensemble of Decision Trees some variance stabilization is gained, but Random Forest thereby loses the property of interpretability.

Gradient Boosting:

Gradient Boosting is another form of an ensemble learner that is based on the weighted combination of weak predictive learners such as Decision Trees, usually

outperforming Random Forest (Hastie et al., 2016). The model is built stepwise and optimized by a differentiably loss function minimizing the in-sample loss (Hastie et al., 2016). It builds the model stepwise, like other boosting methods, and generalizes them by allowing optimization of any differentiable loss function. Whereas in bagging multiple samples of the original training dataset are used to fit a separate decision tree to each one independent from the others and to combine all of trees into a single predictive model, boosting grows the trees sequentially, meaning the information gained from the previous trees is used to grow the current one. This helps to overcome the major issue of training a single large Decision Tree by possibly resulting in an overfitting problem. The gradient boosting algorithm instead learns by constructing a new model based on the previous one and adding the i^{th} base learner $h^{(i)}_{a,t,s}$:

$$\hat{q}^{(i)}_{a,t,s} = \hat{q}^{(i-1)}_{a,t,s} + \lambda_i h^{(i)}_{a,t,s} \tag{32}$$

The model will be improved in such a way that the current residual will be used as an outcome to fit a new Decision Tree and to add this into the originally fitted function with the notion to update the residuals. So, the gradient boosting algorithm fits the new predictor to the residual errors made by the previous predictor. The shrinkage parameter λ_i helps to run the process even slower allowing for more trees and more detailed enhancement of the residuals. Overall in contrary to the bagging methodology, each tree depends on the previous ones (James et al., 2017). Even though the gradient boosting keeps on minimizing the errors, this can cause overfitting in case of a lot of noise in the data and is computationally time and memory expensive, especially because trees are built sequentially (not in parallel as the Random Forest do). Due to the high flexibility, the gradient boosting algorithm also tends to be harder to tune than Random Forest (Hastie et al., 2016). In the analysis part we will be able to assess and improve the goodness-of-fit of the mortality rate estimates produced by standard stochastic mortality models, specifically the ones obtained from the Lee Carter model. The package `gbm` was utilized to implement this algorithm in R (Greenwell et al., 2020). The key parameter configurations were set up following the recommendations of Deprez et al. (2017) and (Levantesi and Pizzorusso, 2019) and will be described in detail in the analysis section.

4.3 Generalized additive model

In recent years, there has been an increasing use of APC methods in various research areas, also due to the availability of more sophisticated statistical approaches. It can be seen as an established tool for the inspection of temporal interrelations, separating cohort effects from age and period (Yang and Land, 2013). All models considered before can be applied for each subpopulation, in this case country and gender only, at least in their original form. Generalized Additive Models (GAMs) with age, period and cohort as possible covariates are used to predict future mortality improvements in a coherent way, possibly allowing for learning from each other both between countries and genders within each country. With the help of tensor products we provide the flexibility to model complex relationships between the covariates and allow intuitively simple models of mortality to be specified. More precisely, age, period (year of death) and cohort (year of birth) come into play as possible

factors to project future mortality building upon the original idea of APC models introduced in Subsection 4.1. Therefore, in this subsection the APC method will be revisited, analysed, refined and showcased in more depth. GAMs can handle it without restrictions and can be seen as extensions of Generalized Linear Models (GLM) resulting in a more robust and flexible method for capturing non-linear effects. GAMs relax the linearity assumption, as this may not apply in practice for complex data structures and therefore replace the simple linear parametric relationships from GLM with smooth functions (Wood, 2017). GAMs can be used for all exponential family responses. Subpopulations are estimated jointly here, additional covariates can be added. The extent to which forecast quality and estimation reliability are improved is discussed in extensive detail in Subsection 5.4.

General GAM Theory:

In the following, a theoretical introduction to the Generalized Additive Regression Model is given as proposed by Hastie and Tibshirani (1987, 1990), Wood (2017). This is then further processed in the context of the APC analysis. It is now possible to consider nonlinear, flexible effects in addition to linear effects for metric covariates. Thus, the relationship of variables, for which the linearity assumption is violated, can be estimated more flexibly. The target variable can be explained by the covariates and an additive error term. So, the assumption is made that the effects of the individual covariates on the target variable are additive. Furthermore, it is assumed that the distribution of the target variable conditional on the covariates belongs to the exponential family. A GAM has the following structure in the context of a simplifying example:

$$g(\mu_i) = \eta_i = \beta_0 + \beta x_{1i} + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) \tag{33}$$

Let x_1, x_2, x_3 and x_4 be metric covariates and y the metric outcome variable. While the effect of x_1 is modelled linearly and is representative for all strictly parametric components of the model, the flexible function $f_2(.)$ allows the non-linear effect of the covariate x_2 to be included, summarizing all linear functionals. Function $f_3(.)$ represents a two-dimensional non-linear interaction between the covariates x_3 and x_4. The log link $g(\mu) = log(\mu)$ is relevant for this thesis as well as the Poisson distribution, which belongs to the exponential family as the distribution of the reponse variable conditional on the covariates. The link function g represents the map from the conditional expectation μ_i of the response to the linear predictor η_i for an observed individual $i \in 1, \ldots, n$ in a dataset. Poisson distribution is defined by the following probability function for $Y_i \sim Poisson(\lambda_i)$ (Fahrmeir et al., 2007):

$$f_{Po}(y_i; \lambda_i) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i\,!} \tag{34}$$

where y_i denotes the number of event occurrences. In this case the link between the linear predictor and conditional expectation is defined in such a way that $\mathbb{E}(y_i|\boldsymbol{x_i}) = g^{-1}(\eta_i) = \lambda_i$, with $\boldsymbol{x_i} = (1, x_{1i}, \ldots, x_{4i})$ for the covariates considered in the model (in the above example 4 covariates) (Fahrmeir et al., 2007). It will be assumed that $y_1, \ldots, y_n$ are samples of independent random variables $Y_1, \ldots, Y_n$

respectively and that Y_i has the probability density (or mass) function of the form of a Poisson distribution and therefore the assumption for the variance $V(y_i|\boldsymbol{x_i}) = \lambda_i$ holds. The flexible functions $f_2(.)$ and $f_3(.)$ of the individual covariates is estimated with the help of the so-called splines, which will be described in the following.

Splines:

Smoothing splines, or splines for short, are used in GAM to maintain a more flexible modelling of the relationships. Visually, it can be thought of as an elastic line that adapts to the data points in a scatter plot. These functions $f_2(.)$ or $f_3(.)$ represent penalised splines based on univariate B-splines (short: *P-splines*) as well as multidimensional P-splines will be introduced.

Univariate splines:

Starting from polynomial splines, which try to represent the data points as precisely as possible by smooth functions, the range of values is divided into $m-1$ intervals, between which the m nodes are defined. After estimating a polynomial of degree l on each interval, the polynomial functions of degree l are formed over $l+2$ knots in order to merge the polynomials of the individual intervals without jumps occurring at the nodes. Finally, the resulting $d = m + l - 1$ basis functions are joined into the a single function. Another feature of the B-Spline base is that the sum of all basis functions at some point $z \in [a, b]$ is $\sum_{j=1}^{d} B_j(z) = 1$.

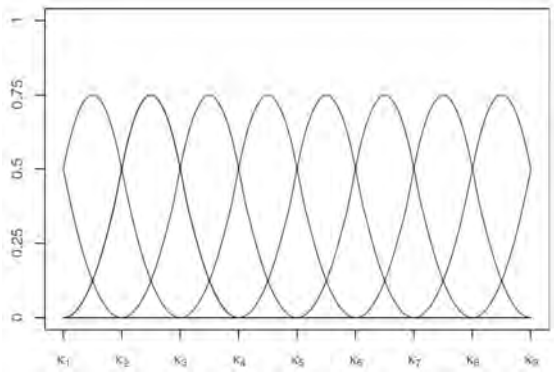


Figure 10: Cubic B-Spline base, $m = 9$ (Fahrmeir et al., 2007)

The figure 10 illustrates the features explained examplarily for a cubis B-spline base over 9 knots. Each smooth function is thus represented by an appropriate basis function for a univariate (equation 29) and a bivariate (equation 30) smooth function, respectively:

$$f_2(x_{2i}) = \sum_{k=1}^{K_2} \gamma_{2,k} B_{2,k}(x_{2i}) \tag{35}$$

$$f_3(x_{3i}, x_{4i}) = \sum_{k=1}^{K_3} \sum_{m=1}^{M_3} \gamma_{2,k,m} B_{3,k}(x_{3i}) B_{3,m}(x_{4i}) \tag{36}$$

In equations (29) and (30) $B_{3,k}$ and $B_{3,m}$ are the B-Splines (Eilers and Marx, 1996), and $\gamma_{2,k}$, $\gamma_{2,k,m}$ the corresponding basis coefficients, estimated from the data. The (marginal) basis dimensions of the p-th smooth are given by K_3 and M_3, respectively.

This results in a continuous function defined on the range of values of the covariates, which is continuously differentiable twice in the respective node. The choice of the number of knots determines to a large extent the smoothness of this function: a high number of knots leads to a rough estimation, whereby the danger of overfitting arises; on the other hand, the fewer knots there are, the smoother the function, which in turn can result in high bias (Fahrmeir et al., 2007). The overarching goal of penalization is to control the tradeoff between bias and variance, ensuring adequate complexity. Besides the number of knots, the flexibility can be regularized by the second–order derivative, i.e. the curvature, used as a measure of variability, so that the following penalisation results for the B-splines just explained, i.e. $\lambda \int f_2''(x_{2i})^2$. Hereby, the smoothness is primarily controlled by the choice of the penalisation or smoothing parameter λ: the larger λ, the smoother the estimate.

Multivariate splines:

It is also possible to consider multidimensional smooth effects. The advantage is that the joint effect of several (usually two) covariates on the target variable is taken into account, like a non-linear, smooth interaction of two continuous covariates. With this kind of interaction, autocorrelation effects can be treated, since, for example, two covariates are simultaneously included in the estimation of the splines (Wood, 2017).

Parameter estimation:

The base expansion of the smooth functions allows the resulting model to be linear in the parameters and thus to estimate this as a standard generalized linear model (GLM), e.g., by Fisher scoring. However, the basis dimensions K_3 and M_3, and thus the estimation, are somewhat arbitrary. Therefore the models tend to overfitting, especially for large K_3 and M_3, respectively. For this reason, a quadratic penalty of the base coefficients is used for each smooth function, as explained below. The upper expression for smooth function can be rewritten in the general case of the p-covariate into this form: $f_p(\boldsymbol{x_p}) = \sum_{k=1}^{K_p} \gamma_{p,k} B_{p,k}(\boldsymbol{x_p}) = \boldsymbol{B_p}\boldsymbol{\gamma_p}$, so that $\boldsymbol{B_p}$ represents the design matrix, $\boldsymbol{x_p}$ the observed values and $\boldsymbol{\gamma_p}$ the vector of basis coefficients of the p-th smooth function. After selecting $\tilde{\boldsymbol{S}}_{\boldsymbol{p}}$ according to the smooth type, quadratic penalty of the coefficients of each smooth function will result in $\boldsymbol{\gamma_p'}\tilde{\boldsymbol{S}}_{\boldsymbol{p}}\boldsymbol{\gamma_p}$. For the total penalty over all smooth functions $\tilde{\boldsymbol{S}}_{\boldsymbol{p}}$ will be padded with zeros to make sure that the contribution of other coefficients is set to zero, yielding in $\boldsymbol{S_\lambda} = \sum_{p=1}^{P} \lambda_p \boldsymbol{\gamma' S_p \gamma}$ (Wood, 2017). Hereby, λ_p represents the smoothing parameter and $\boldsymbol{S_p} = \boldsymbol{D_2'D_2}$ is a diagonal block matrix resulting from the second order differences, defined as:

$$\boldsymbol{D_2} = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{pmatrix}$$

so that $\lambda_p \boldsymbol{\gamma}' \boldsymbol{S_p} \boldsymbol{\gamma}$ captures the penalty for the smooth function f_p. While $\boldsymbol{\gamma} = (\boldsymbol{\beta}', \boldsymbol{\gamma_1'}, \ldots, \boldsymbol{\gamma_P'})'$ represents the coefficient vector, the λ_p stand for the smoothing parameters of the p-th smooth function to control the smoothness of the estimated relationship: for higher λ_p values, the f_p approaches a constant zero function, in contrary for smaller values f_p becomes unpenalized and thus more wiggly. Paraphrasing the original model with the upper notation, the log-likelihood $l(\boldsymbol{\gamma})$ will be penalized resulting in the following expression of penalized log-likelihood:

$$l(\boldsymbol{\gamma}) - \frac{1}{2} \sum_{p=1}^{P} \lambda_p \boldsymbol{\gamma}' \boldsymbol{S_p} \boldsymbol{\gamma} \tag{37}$$

for the following model:

$$g(\mu_i) = \boldsymbol{X_i} \boldsymbol{\gamma}, y_i \sim Po(\mu_i) \tag{38}$$

The optimization involves two dimensions: estimation of coefficient vector $\boldsymbol{\gamma}$ and the estimation of the smoothing parameters $\boldsymbol{\lambda}$. For fixed smoothing parameters, the coefficients $\boldsymbol{\gamma}$ can be estimated by the penalized, iteratively re-weighted least squares (P-IRLS) algorithm explained step by step below (Wood, 2017):

1. Initialize expectation $\hat{\mu}_i = y_i$ and linear predictor $\hat{\eta}_i = g(\hat{\mu}_i)$
2. Calculate: $\tilde{y}_i = g'(\hat{\mu}_i)(y_i - \hat{\mu}_i)/\alpha(\hat{\mu}_i) + \hat{\eta}_i$ and $w_i = \alpha(\hat{\mu}_i)/(g'(\hat{\mu}_i)^2 V(\hat{\mu}_i))$, here $V(\hat{\mu}_i)$ is the family specific variance function. Under Poisson distribution assumption $V(\mu_i) = \mu_i$. Let further $\alpha(\mu_i) = (1 + (y_i - \mu_i)(V'(\mu_i)/V(\mu_i) + g''(\mu_i)/g'(\mu_i)))$
3. Minimize and obtain the P-IRLS estimate $\hat{\boldsymbol{\gamma}}$ from a $\underset{\boldsymbol{\gamma}}{\operatorname{argmin}}(\tilde{\boldsymbol{y}} - \boldsymbol{X}\boldsymbol{\gamma})'\boldsymbol{W}(\tilde{\boldsymbol{y}} - \boldsymbol{X}\boldsymbol{\gamma}) + \sum_{p=1}^{P} \lambda_p \boldsymbol{\gamma}' \boldsymbol{S_p} \boldsymbol{\gamma}$, where $\tilde{\boldsymbol{y}} = \tilde{y}_i, \ldots, \tilde{y}_n$ is the vector of pseudo observation $\tilde{y}_i$ calculated in the previous step and $\boldsymbol{W}$ is a diagonal $n \times n$ matrix with elements $\boldsymbol{W}_{i,i} = w_i$
4. Update $\hat{\boldsymbol{\eta}} = \boldsymbol{X}\hat{\boldsymbol{\gamma}}$
5. Iterate $2 - 4$ until convergence

In other words, a maximization of the penalized likelihood is performed. The sum of the penalized weighted least squares is minimized and in each step new data and corresponding weights are created on the basis of the previously estimated model. After that new coefficients are estimated by the minimization of the penalized least squares method. After that, pseudo data are created again. The same process is repeated until a pre-specified convergence criterion is met. The coefficients in the last step represent the final parameter estimation in the respective model.

The estimation method for the penalty parameter used in the models is the so-called REML-based approach, which is derived from Wood (2017). Here, the smoothing parameters are considered to be part of the variance components of

random effects and therefore they can be estimated by variance estimation using Restricted Maximum Likelihood. Let $\boldsymbol{\gamma}$ be the random effects with multivariate normal distribution $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{S}_{\boldsymbol{\lambda}}^{-})$, where $\boldsymbol{S}_{\boldsymbol{\lambda}}^{-}$ is a generalized inverse of $\boldsymbol{S}_{\boldsymbol{\lambda}}$ with

$$f(\boldsymbol{\gamma}) = \frac{|\boldsymbol{S}_{\boldsymbol{\lambda}}|_{+}^{1/2}}{\sqrt{(2\pi)^{dim(\boldsymbol{\gamma})-M}}} \exp\left(\frac{-\boldsymbol{\gamma}'\boldsymbol{S}_{\boldsymbol{\lambda}}\boldsymbol{\gamma}}{2}\right) \tag{39}$$

The above equation consists of a generalized determinant $|\boldsymbol{S}_{\boldsymbol{\lambda}}|_{+}$ and the dimension M of the null space of $\boldsymbol{S}_{\boldsymbol{\lambda}}$. By integrating, the restricted likelihood can be derived as follows:

$$\begin{aligned}
\int f(\boldsymbol{y}, \boldsymbol{\gamma}) d\gamma &\approx \int exp\left(\log(f(\boldsymbol{y}, \hat{\boldsymbol{\gamma}})) + \frac{1}{2}(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})' \frac{\partial^2 \log(f(\boldsymbol{y}, \boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} (\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})\right) \partial \boldsymbol{\gamma} \\
&= f(\boldsymbol{y}|\hat{\boldsymbol{\gamma}}) f(\hat{\boldsymbol{\gamma}}) \\
&\int \exp\left(\frac{1}{2}(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})' \frac{\partial^2 \log(f(\boldsymbol{y}|\boldsymbol{\gamma})) + \log(f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} (\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})\right) \partial \boldsymbol{\gamma} \\
&= L(\hat{\boldsymbol{\gamma}}) f_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\gamma}}) \int \exp\left(-\frac{1}{2}(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})'(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} + \boldsymbol{S}_{\boldsymbol{\lambda}})(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})\right) \partial \boldsymbol{\gamma} \\
&= L(\hat{\boldsymbol{\gamma}}) f_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\gamma}}) \frac{\sqrt{(2\pi)^{dim(\boldsymbol{\gamma})}}}{|\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} + \boldsymbol{S}_{\boldsymbol{\lambda}}|^{1/2}}
\end{aligned} \tag{40}$$

The Taylor expansion around $\hat{\boldsymbol{\gamma}}$ is responsible for the second term, whereas the subsequent equation comes up by using $f(\boldsymbol{y}, \boldsymbol{\gamma}) = f(\boldsymbol{y}|\boldsymbol{\gamma}) f(\boldsymbol{\gamma})$ twice. The negative, weighted Hessian matrix is given by $\frac{\partial^2 l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}' \partial \boldsymbol{\gamma}} = -\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}$ with the diagonal matrix of Newton weigths $\boldsymbol{W}$ obtained from the P-IRLS step conditional on the current values of $\boldsymbol{\lambda}$. The last step is achieved by the integration of a multivariate normal (Wood, 2017, Bender, 2018).

The application of logarithm results in the final REML criterion suggested by Wood (2011):

$$\mathcal{V}(\boldsymbol{\lambda}) \approx l(\hat{\boldsymbol{\gamma}}) - \frac{\hat{\boldsymbol{\gamma}}'\boldsymbol{S}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\gamma}}}{2} - \frac{\log|\boldsymbol{S}_{\boldsymbol{\lambda}}|_{+}}{2} - \frac{\log|\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} + \boldsymbol{S}_{\boldsymbol{\lambda}}|}{2} + \frac{M}{2}\log 2\pi \tag{41}$$

Finally, according to Wood (2011), to guarantee convergence, an outer iteration will be used to estimate the smoothing parameters by optimizing the REML criterion, to be followed by an inner P-IRLS iteration (Wood, 2017) in order to obtain $\hat{\boldsymbol{\gamma}}$ conditioned on the updated value of $\boldsymbol{\lambda}$. The two steps are then repeated until convergence is met. As Wood (2017) shows, this approach is preferable to the Generalized Cross Validation (GCV) criterion. The latter is a method for choosing the optimal effective degrees of freedom (edf) based on a trade-off between bias and variance for the smoothing procedure. This method is implemented in the package `mgcv` in the function `gam` as `method = REML` (Wood, 2021).

APC context and identification problem:

The usage of age-period-cohort (APC) analysis framework is supposed to give insights into the extent to which age, period and birth cohorts are related to the human mortality. According to Yang and Land (2013), whereas age effects are associated with the ageing process of an individual and period effects are connected to external events and environmental changes, the cohort effects represent the specific groups who experience the same events in a specific time frame as well as benefit of medical and technological improvements of that time. Therefore, the consideration of these three dimensions to explain temporal trends is recommended by research (Oppermann, 1995). Since each of the three components of the temporal structures is a linear combination of the other two (i.e. $age + cohort = period$), identification problems arise. This leads to the fact that the effects cannot be clearly assigned to the individual components and thus cannot be unambiguously interpreted. In a classical regression framework this leads to collinearity problems and the model is then not able to identify distinct effects of these components (Clayton and Schifflers, 1987). With the help of GAMs and bivariate tensor product splines between age and period the identification issue is thus solved without using restrictive assumptions (in form of constraints) on the time-related effects. This happens, since the cohorts are modeled implicitely, as an interaction between age and period. This approach is applicable for both aggregated and individual data. Moreover, we leverage the fact that the marginal effect of one of the three features is automatically part of the interaction space of the two others. Thus this approach allows to get insights into the structure of the available data and to reasonably and systematically analyse them. As mentioned earlier the multiplicative three-factor regression modelling approach is not applicable because of the linear dependency of age, period and cohort. Contraints have to be set up, more or less the same as seen in Subsection 4.1. However, Clements et al. (2005) suggests using APC model in GAM framework using bivariate spline function depending on age and period resulting in a two-dimensional interaction surface. This implicitely contains the cohorts on its diagonals. In the context of this section, the index for time in years is changed from time $t \in T$ to period $p \in P$ to keep consistency of terminology. The number of deaths is still assumed to follow a Poisson distribution taking the form $D_i \sim Poisson(E_i \cdot \mu_i)$ for all $i = 1, \ldots, n$, with $n = (Age \times Period \times Country \times Gender)$ indicating the data dimension. For a fixed year, the number of deaths for age a in a portfolio are assumed to be i.i.d.. For aggregated data the model structure in its general form is then given by:

$$g(\mu_{apc,i}) = \beta_0 + f_{ap}(age_i, period_i) + g(E_i) \tag{42}$$

with f_{ap} representing a non-linear interaction surface with a two-dimensional spline basis, in this case a tensor product basis which is defined as the Kronecker product of two one-dimensional marginal spline basis over age and period. Thereby, each marginal basis function of age is multiplied with each marginal basis function of period in order to maintain the two-dimensional spline basis. In this case, penalized B-splines (Eilers and Marx, 1996) are used to define marginal spline bases, each of them consisting of ten basis functions. The penalization is done by second-order differences. Even if the non-linear relationships can be modeled accurately, the use of penalty integrated both in the univariate and bivariate P-splines can avoid overfitting

(Weigert et al., 2021, Wood, 2017). More precisely, a semiparametric additiv Poisson regression with log link function $\log(\mu_i) = \eta_i$ will be applied with offset for the exposure $\log(E_i)$ to model the aggregated death counts and thus mortality rates. This yields in the following model specification:

$$\log(\mu_{apc,i}) = \beta_0 + f_{ap}(age_i, period_i) + \log(E_i) \tag{43}$$

This model is referred to as the pure APC model. In contrast, the coherent APC model captures the subpopulations discribed in the previous subsection explicitely, by estimating a bivariate function of age and period for each country and each gender in form of an interaction for the tensor product:

$$\log(\mu_{apc,i}) = \beta_0 + f_{ap,s_i}(age_i, period_i) + \log(E_i) \tag{44}$$

In order to visualize the marginal effects of each component in an accessible way the temporal developments will be condensed in one specific dimension only, to be averaged over the respective other component, whereas the cohort values can be viewed as post-stratification (Weigert et al., 2021). In the case of interactions, geometric-averaged statements about the different subpopulations can also be made. The tensor product estimates provide the required information and will be averaged in the following way:

$$\begin{aligned} f_a(age_i) &= \frac{1}{|P|} \sum_{period_i \in P} f_{ap}(period_i | age_i), \\ f_p(period_i) &= \frac{1}{|A|} \sum_{age_i \in A} f_{ap}(age_i | period_i), \\ f_c(cohort_i) &= \frac{1}{|A| \cdot |P|} \sum_{age_i \in A} \sum_{period_i \in P} f_{ap}(age_i, period_i | cohort_i). \end{aligned} \tag{45}$$

The inclusion of further variables potentially associated with human mortality is considered in the GAM framework. The model specification and thus the linear predictor will then be extendend:

$$\begin{aligned} \log(\mu_{apc,i}) = \beta_0 + f_{ap,s_i}(age_i, period_i) + \beta_{bmi} bmi_i + \beta_{unemployment} unemployment_i \\ + \beta_{tax} tax_i + \beta_{alcohol} alcohol_i + \log(E_i) \end{aligned} \tag{46}$$

The suitability for the modelling of socio-economic mortality differentials and interpretation of the linear effects in the context of aggregated data will be discussed in detail in Subsection 5.4.

4.4 Trend forecast considering COVID-19

The GAM with P-Splines and Tensorproduct splines extrapolates future mortality rates. The shape of the curve depends primarily on how many basis functions are used. Wood (2017) suggests to be careful with extrapolation in such non-linear estimates and not to project too far into the future. In the case of the year 2020, especially for higher ages, an outlier can be detected as indicated in Section 2 and will be discussed in more detail later. To consider a kind of structural change at the edge of a time series requires a special handling during the prediction process. We have developed different scenarios to estimate the development of mortality rates in the upcoming years. Since at the time of writing this thesis an immense uncertainty both in epidemiological and medical terms is associated with COVID-19 pandemic, the conclusions that are drawn from this must be taken with all caution. All scenarios are based on the core model defined in equation 43. As the aim now is to train the model beyond 2015 which was the upper limit of the training in the context of the study of first research question, we additionally include a covariate indicating the year 2020 with special aspect of COVID-19 yielding in the following model definition:

$$\log(\mu_{apc,i}) = \beta_0 + f_{ap,s_i}(age_i, period_i) + \beta_{covid,s_i} covid_i * s_i + \log(E_i) \tag{47}$$

The asterix in the model formula specifies an interaction between the subpopulation s_i and the COVID-indicator $covid_i$ defined as:

$$covid_i = \begin{cases} 1, & \text{for } period_i = 2020 \\ 0, & \text{for } period_i \leqslant 2019 \end{cases}$$

The consideration of adding this interaction has its origin in the intention to represent gender-specific country effects properly. Cultural behavioral changes on the one hand, political-social decisions in the course of COVID-related events with correspondingly different consequences on the other hand justify this move. In the following, four possible scenarios and the prediction methodology behind them are described:

Scenario I: COVID-19 will disappear in the future.

COVID-19 is a residual and special event and does not affect mortality in future years. Thus, excess mortality averages out over the coming years and it is assumed that there are no long-term effects due to COVID. Technically, this can be implemented by training the model with data up to 2019 only and then predict for 2020-2025.

Scenario II: Expect full COVID-effect in the future.

It is probably a demographically dramatic development scenario. The COVID-situation will continue just as it did in 2020. In this case, the model is trained with original mortality data up to and including the crucial year 2020 and for the next

years, assuming that the indicator variable is still set to 1.

Scenario III: Flattening COVID-effect over years.

This is similar to Scenario II, except that we do not assume a constant effect in the future. The next years are influenced by the residual event. There are long-term effects due to the pandemic, but their influence decreases exponentially.

Scenario IV: Adjustment for excess mortality.

Here, COVID is considered as a residual or special event without influencing mortality in the upcoming years, as in Scenario I. The only difference in this scenario is that excess mortality does not average out over the next few years and therefore must be explicitly considered. Baseline mortality remains the same, so there are no behavioral changes in all age groups due to the pandemic. We model this by calculating the excess mortality for the year 2020 in the first step. To do this, the expected death numbers are used based on the model that was fitted for Scenario I. These counts are compared to the actual mortality counts. The positive difference represents then the excess mortality. The population for 2021 is thus adjusted so that the age- and cohort-specific population from 2019 is reduced by the actual deaths in 2019 and those in 2020. For newborns, the previous year's information is transformed unchanged, as COVID is expected to have little to no impact on this age group. The model is ultimately trained with the expected death counts for 2020 and 2021, and with the adjusted population count for 2021.

As mentioned above, all scenarios are subject to considerable uncertainties. The results provided as well as the assumptions and limitations are discussed in detail in Subsection 5.5.

4.5 Goodness-of-fit evaluation

The inspection of the standardised deviance residuals of the fitted model or subpopulation is a common way of assessing the goodness-of-fit of the models. With a Poisson component, it is appropriate to look at the scaled variance residuals in each case conditioned on the subpopulation s, which are denoted as:

$$D(d_{a,t,s}, \hat{d}_{a,t,s}) = \sum_s \sum_a \sum_t w_{a,t,s} dev(a,t,s) \tag{48}$$

$$r_{a,t,s} = sign(d_{a,t,s} - \hat{d}_{a,t,s}) \sqrt{\frac{dev(a,t,s)}{\hat{\phi}}}, \text{with } \hat{\phi} = \frac{D(d_{a,t,s}, \hat{d}_{a,t,s})}{K - \lambda} \tag{49}$$

Hereby, $dev(a,t,s) = 2d_{a,t,s} \log(\frac{d_{a,t,s}}{\hat{d}_{a,t,s}}) - (d_{a,t,s} - \hat{d}_{a,t,s})$ is the total deviance of the model, where $K = \sum_{s \in S} \sum_{a \in A} \sum_{t \in T} w_{a,t,s}$ indicates the number of observations in the data and λ is the effective number of parameters in the model.

Regular patterns in the residuals indicate the model's inability to adequately describe all features in the data (Villegas, 2015). In order to evaluate the accuracy of the model predictions both in-sample and out-of-sample back-testing is used. For that, the Root Mean Square Error (RMSE) is used, which quantifies the difference

between the mortality rates predicted by the model and the actual mortality rates observed. In Section 5 the goodness-of-fit analysis will be presented in tabular or graphical form, sometimes offering threefold inspection of error behaviour from each age, cohort and period perspective. The RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum_{s \in S} \sum_{a \in A} \sum_{t \in T} (d_{a,t,s} - D_{a,t,s})^2}{n}} \tag{50}$$

with $n = (A \times T \times S)$ as data dimension, $d_{a,t,s}$ as the expected number of deaths and $D_{a,t,s}$ as the actual number of deaths observed at age a in year t. In Section 5 for each model class the accuracy of the predictions were assessed over various intervals, depending on either the data availability or the structural changepoint detection.

Alternative to RMSE, the Mean Absolute Percentage Error (MAPE) is proposed in some places:

$$MAPE = \frac{100}{n} \sum_{s \in S} \sum_{a \in A} \sum_{t \in T} \left| \frac{D_{a,t,s} - d_{a,t,s}}{D_{a,t,s}} \right| \tag{51}$$

Its advantage is the relative proportion to the actual mortality rates, so that, for example, smaller mortality rates at younger age categories are equally included and considered in the calculation (Villegas, 2015). For projections into the future no evaluation is possible by nature.

5 Results

The analyses were conducted based on the theoretical background presented in the previous sections. Before the first evaluation results are introduced in this section, we outline the evaluation plan. The procedure is twofold. Initially, we address the first research question investigating how the state-of-the-art mortality models can be improved with respect to fit and forecast. This includes the evaluation on the contribution of machine learning methods in mortality modeling. Apart from that, the aim is to find out to what extent the integration of further socio-economic, health-risking variables is beneficial in this context.

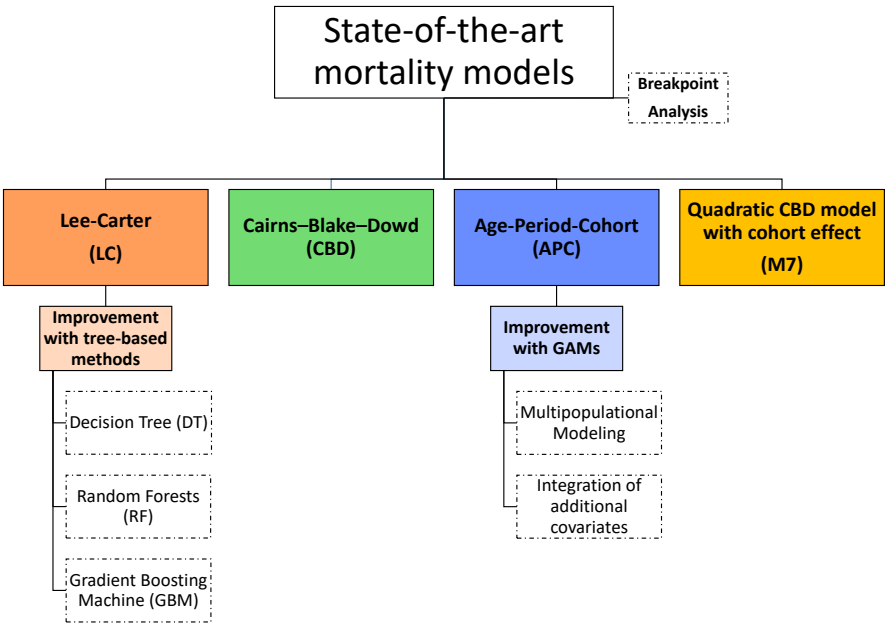


Figure 11: Schematic proceeding

Figure 11 provides an overview of the schematic proceeding for the first part of analysis involving the application of the classical models. This is primarily the Lee Carter model under Poisson assumption introduced in Section 2. For comparison, three further extensions of the LC model are applied to the present data: CBD model (Cairns et al., 2006), Quadratic CBD model with cohort effects (M7) (Cairns et al., 2009) and Age-Period-Cohort Model (APC) (Clayton and Schifflers, 1987, Hobcraft et al., 1982, Holford, 1983, Currie, 2006). A breakpoint analysis is performed, after the first insights and characteristics of each model are obtained. This happens for each subpopulation, that is, for each combination of country and gender. Without anticipating too much, it is already warned at this point that not every subpopulation has a break in the time series. The analysis allows us to establish a preference for two models, LC and APC. This finding suggests to use these two models as a foundation for further development and examination. With the machine learning methods, we revisit the Lee Carter model and hope to improve upon the methodology described in Subsection 4.2. Specifically, the three tree-based models are used here: Decision Tree, Random Forests, Gradient Boosting. The thorough evaluation of the improvement in model performance contains two aspects, on the one hand an improvement in goodness of fit and on the other hand an improvement in goodness of forecast must be achieved. The extent to which this contribution is beneficial and suitable in the case of available data is discussed in detail in Subsection 5.3.
The basic idea of the APC model is exploited to enable further modeling, so that the idea to model mortality rates capturing the three important components age, period and cohort is transferred to the GAM universe. The primary advantages here are that no additional constraints are necessary and a joint modeling of all subpopulations is guaranteed. Furthermore, this allows the addition of further covariates in a fairly convenient manner. This approach is devoted to examining the trends and financial implications of socio-economic and health disparities in mortality.
The second part of the evaluation discusses the question of how mortality will develop in the upcoming years in different countries. As there is a kind of changepoint at

the boundary of the time series, which is related to the COVID-19 situation in 2020, standard breakpoint analysis cannot be applied. Thus, the effect of COVID-19 must be treated specially and with all caution.
At the time of writing this thesis, both the medical and epidemiological situation about the future is uncertain. In order to be able to model and forecast into the future, four scenarios are set up. These are of course subject to certain assumptions, which in turn require an explicit investigation. Therefore, the discussion of these assumptions and scenarios for the future situation is an essential part of the content. After all, this will enable the forecasts of mortality rates to be presented as applied cases. Recently, both mortality rates and life expectancy have emerged as important metrics and headlines for various research projects and newspaper articles. To accommodate this, based on our results, another focus will be to convert the predictions of mortality rates into life expectancy and then make them available in tabular form as a reference tool especially for older age categories and for the years 2022 and 2023. Even though the aim of outlining the potential impact of the COVID-19 pandemic is of great importance for modeling mortality, at this time, however, statements and analyses of the impact of COVID-19 are highly speculative in nature.
The data set, enriched according to the procedure described in Section 3, provides a solid basis for the analyses. First, the mortality data for up to and including the year 2020 are now available, for all countries Finland, Germany, Italy, the Netherlands and the United States. The results are presented with particular reference to the US and German populations, to avoid exceeding the volume of the thesis in some places. The age categories 0-90 were taken into account, excluding the ages above 90, in order not to provoke distortions due to insufficient data in the even higher categories. Finally, the data are enriched by additional variables that could be relevant for differentiating mortality. These are: Alcohol consumption, proportion of overweight, tax ratio, unemployment rates.
With this foundations, we move now to the results subsections.

5.1 State-of-the-art mortality models

We first apply the Lee Carter (LC) model to all subpopulations. The fitting procedure is done for the years from 1950 up to including 2010. Germany is an exception, with a rather small data base. Here, the years 1990-2010 are used for fitting. Data availability is related to reunification. We refrain from merging the mortality data of East and West Germany for the years before, because the data collection strategies do not match exactly (HMD, 2021). The years 2011-2019 are used for evaluating the goodness of forecast. The year 2020 is intentionally omitted from the testing procedure for the time being due to its special outlier character. For fixed gender and country the parameters $\beta_{a,s}$, $\alpha_{a,s}$ and $\kappa_{t,s}$ (see Subsection 4.1) are fitted to the mortality data using `StMoMo` R-package (Villegas et al., 2016). The implementation comprises the exposures $E_{a,t,s}$ and the numbers of deaths $D_{a,t,s}$ as inputs. The LC mortality rates are then fitted to the data for each subpopulation and time range.

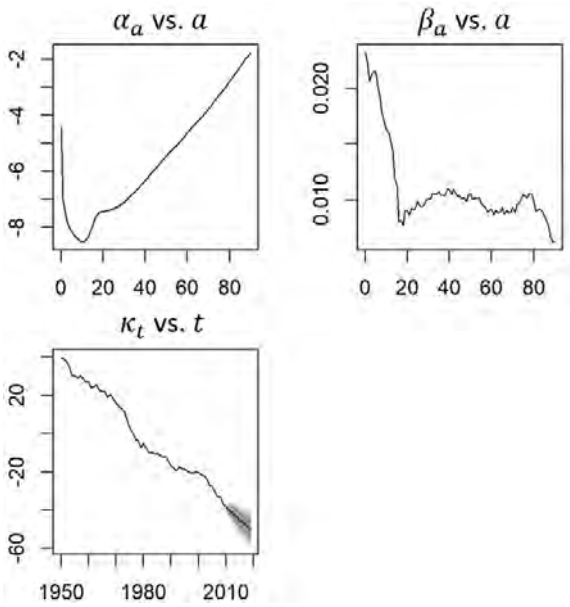


Figure 12: Parameters for the Lee-Carter (LC) model fitted to the US female population for ages 0-90 and the period 1950-2010. The period effect is projected up to year 2019.

Figure 12 provides a visualization of fitted parameters of the Lee Carter model for the US female population. The course of the period effect $\kappa_{t,s}$ confirms and captures the initial statement that over the past 50 to 60 years mortality rates have been decreased steadily, in this particular case even fairly linearly on log scale. The latter fact is useful in terms of forecasting a linear trend, which is indicated by the greyish part on the right hand side of the graph and will be discussed later. This has led to an increase in life expectancy. To illustrate the inspiration of this parameter combination, the figure shows that the decline in mortality rates over the years is not the same in all age categories.

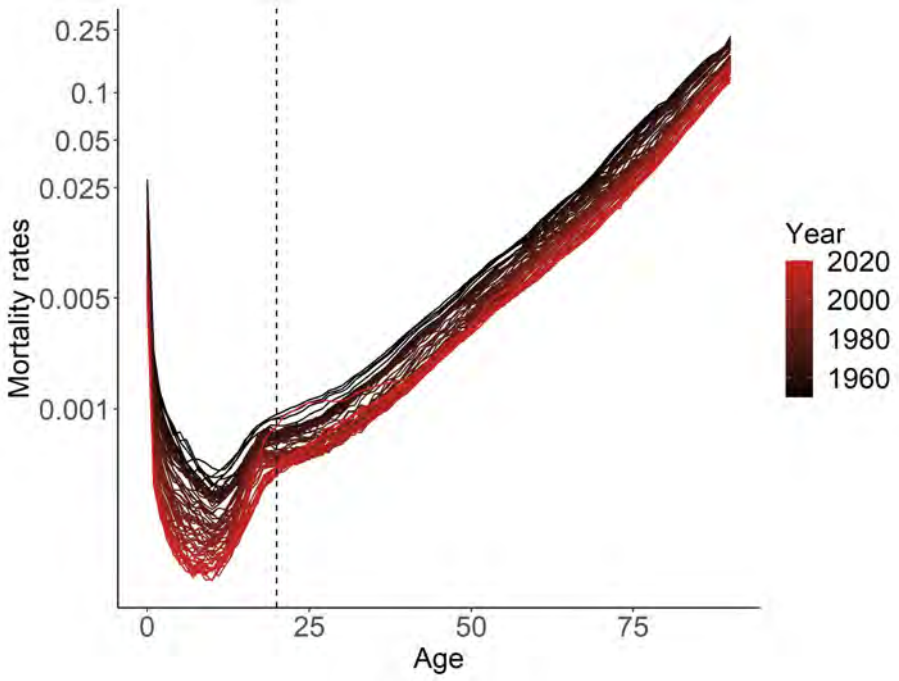


Figure 13: Mortality rates for the US female population aged 0-90 years. The color scheme shows the evolution over the years 1950-2020. The dotted horizontal line marks the age of 20. Vertical axis spacing represents mortality rates with ticks on a logarithmic scale.

The typical course for each year is obviously similar: mortality is quite high for infants and then gradually decreases until it reaches a minimal risk at about 12 years of age. In adolescence, mortality increases abruptly until the age of 20. Thereafter, it increases exponentially until the highest age. This general progression of the curves, if one were to imagine them averaged, without taking into account individual yearly differences, defines the overall trend of the age effect and comprises the static part $\alpha_{a,s}$, capturing the total shape of the mortality by age: Mortality decreases when females in the US are younger, then it goes up and afterwards it increases even more, behaving nearly linearly on logarithmic scale (figure 12).

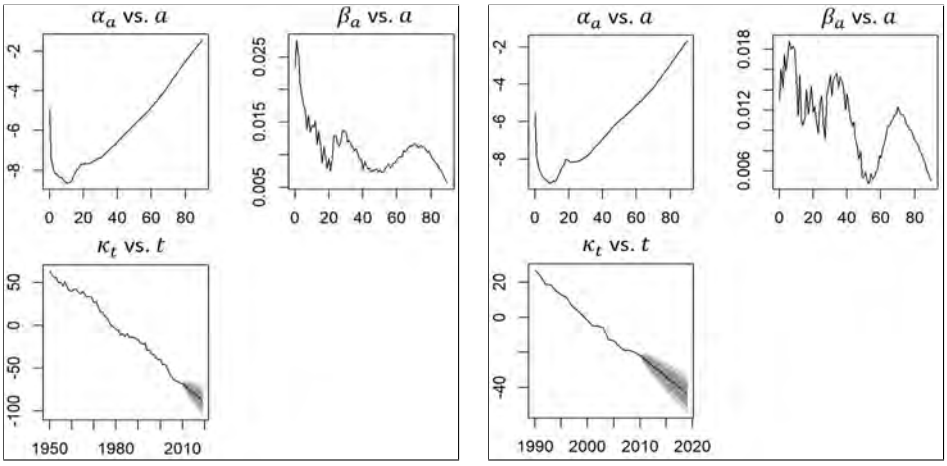
Another question is how the mortality by age evolves in time. To understand the overall trend of time effect the colour gradient is decisive, from dark to light, that is, from earlier to more recent years, the mortality rate decreases for all age categories as the lines take lower levels. This is in line with $\kappa_{t,s}$ capturing the period effect, so the decreasing mortality in time.

Finally, the improvement rate $\beta_{a,s}$ is about how much improvement there is over the years per age category. Visually, this can be seen by the range of the lines, which differs according to age. Looking at the bunch of lines and paying attention to the color scheme, one can see that the decline is greater in some age groups and less in others, depending on the year.

Figure 12 shows for example, that the estimated $\beta_{a,s}$ parameters for age categories younger than 20 are more affected by the period change than the higher ages. This is directly related to the observed mortalities, because the greater the differences between the lines, the greater the annual improvement (in figure 13). To summarize, $\beta_{a,s}$ analyzes at which ages mortality has improved faster. Thus, in this case, it can

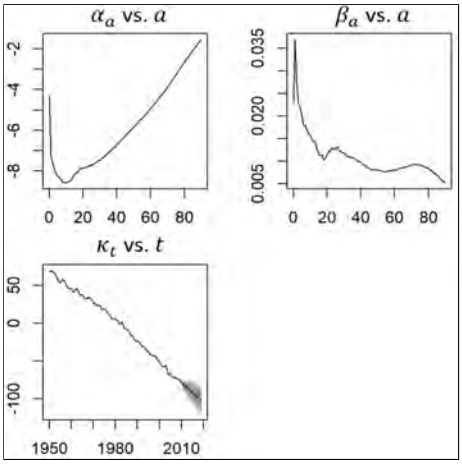
be said that the improvement in mortality for the 10-year-old girls was faster than that for the 60-year-old women in US.

This evolution of mortality rates is what the Lee Carter, as a well-known, widespread and used method, achieves to model, by letting the log of mortality rates depend on age and time given in years.

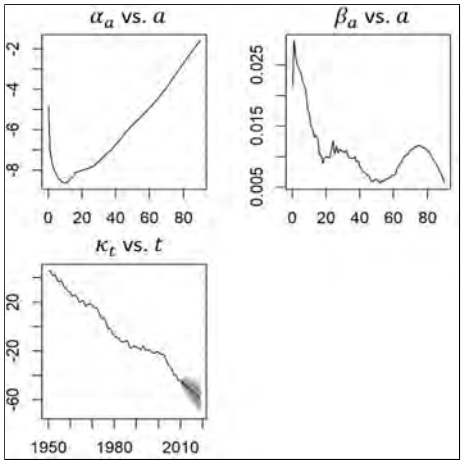


(a) Finnish females

(b) German females



(c) Italian females



(d) Dutch females

Figure 14: Parameters for the Lee-Carter (LC) model fitted to the Finnish, German, Italian and Dutch female populations for ages 0-90 and the period 1950-2010 (1990-2010 for Germans). The period effect is projected up to year 2019.

For female populations in other countries, the estimates of the static age effect $\alpha_{a,s}$ as well as those of the period effect look similar. There are striking differences in the age-specific speed of mortality improvement. In the Netherlands, Germany, and Finland, women aged 60 and older have experienced a comparatively large decline in mortality rates over the years. Moreover, among German women in general, a high variability is present in the course of the estimation of the $\beta_{a,s}$ coefficient. However, while the model works reasonably well, with the limitation that the age-specific mortality improvement rates $\beta_{a,s}$ are assumed to be constant over time. The assumption seems to have been violated in several low-mortality countries. Research states that rates of mortality improvements have tended to decline over time at

younger ages, and they have risen at older ages (Kannisto et al., 1994, Vaupel et al., 1998, Wilmoth and Horiuchi, 1990). Some of extensions of the Lee Carter model, as explained in detail in Subsection 4.1, introduce the cohort effect additionally which is an important component that make mortality explicitly dependent also on the time people were born, as for example in the APC or M7 model. Also, other authors have chosen to introduce more age-period effects, such as for CBD and M7 models. It is perhaps not surprising that people who live in different years or generations have different mortality experiences. There is much research particularly for the British population, but the findings there can be generalized. For example, Willets (2004) finds evidence for the existence of cohort trends in the population of England and Wales. Murphy (2009) discusses the "golden generations" of the U.K. population born in the early 1930s that experienced exceptionally rapid improvement in mortality rates. Willets (2004), Murphy (2009), and Murphy (2010) discuss cohort effects in detail, including their identification using mortality data, and competing explanations. The aforementioned studies, however, are not model-based but rely on empirical data analysis and qualitative analyses such as descriptive and graphical representations to point out the importance of cohort effects on population mortality. The graphics in figure 15 show exemplary the estimated parameters for the APC model, also again for the women in the US.

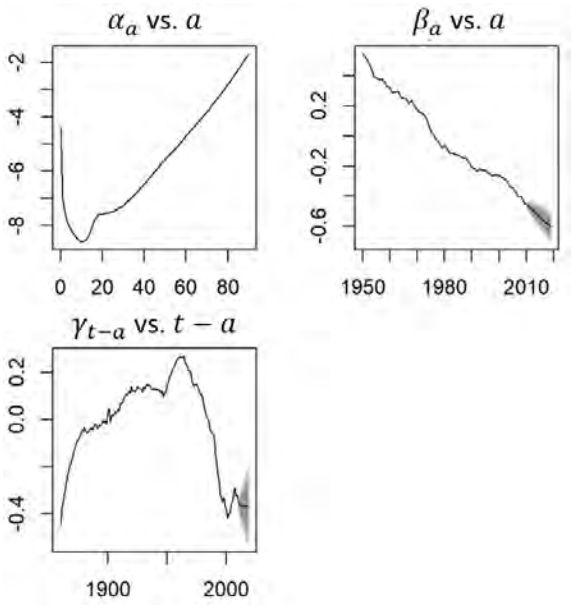
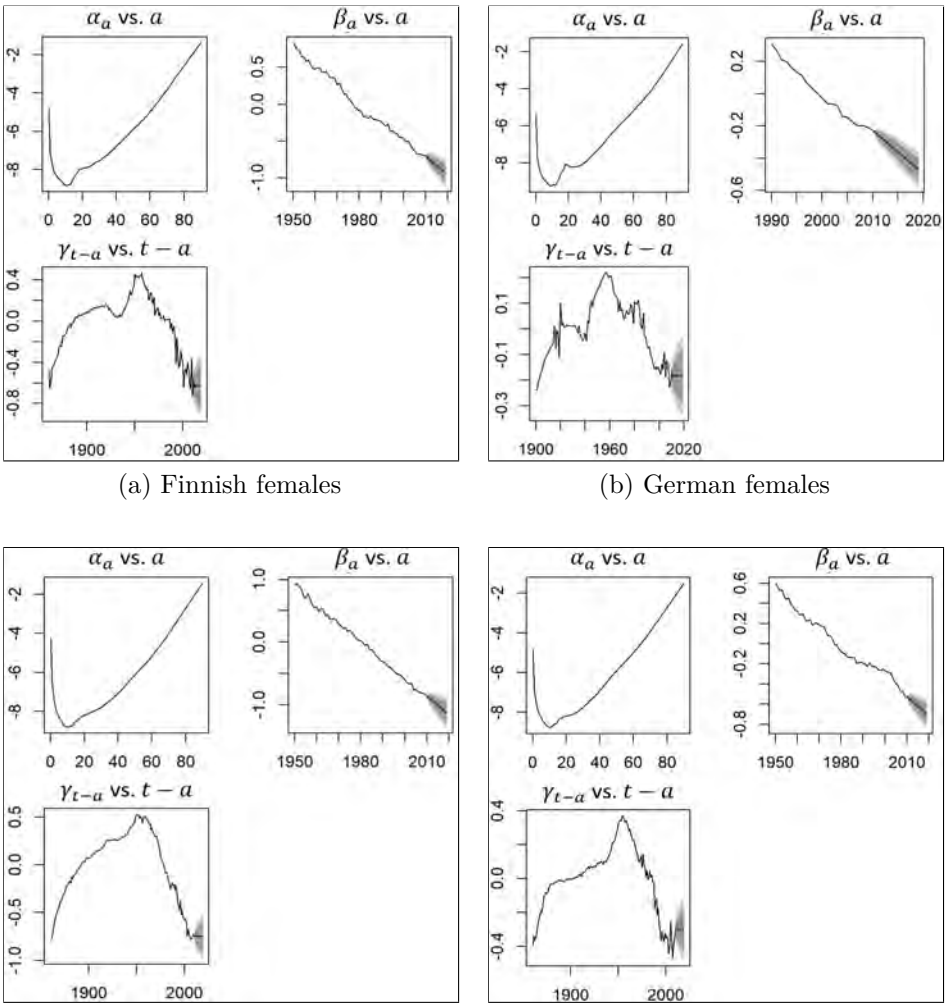


Figure 15: Parameters for the Age-Period-Cohort (APC) model fitted to the US female population for ages 0-90 and the period 1950-2010. The period effect is projected up to year 2019.

The estimated parameters for the general relationship between age and mortality as well as period and mortality remain broadly similar to the estimates from the Lee Carter model. For people born until 1960, the rough effect of mortality is upward. In contrast, for those born between 1960 and 2000 the effect decreases and increases slightly thereafter. Recall that according to the model equation discussed in the Section 4, this is an additive effect on log mortality. For other countries the cohort effect estimation is quite similar except for the small increase in the last several years, as shown in figure 16. For German females, the main course is similar, but

with more fluctuations. The predictions for the $\gamma_{t-a,s}$ coefficient is nearly constant until 2020, which holds for all subpopulations. However, due to the absence of $\beta_{a,s}$ coefficient, the APC model is inable to allow for varying improvement rates with age.

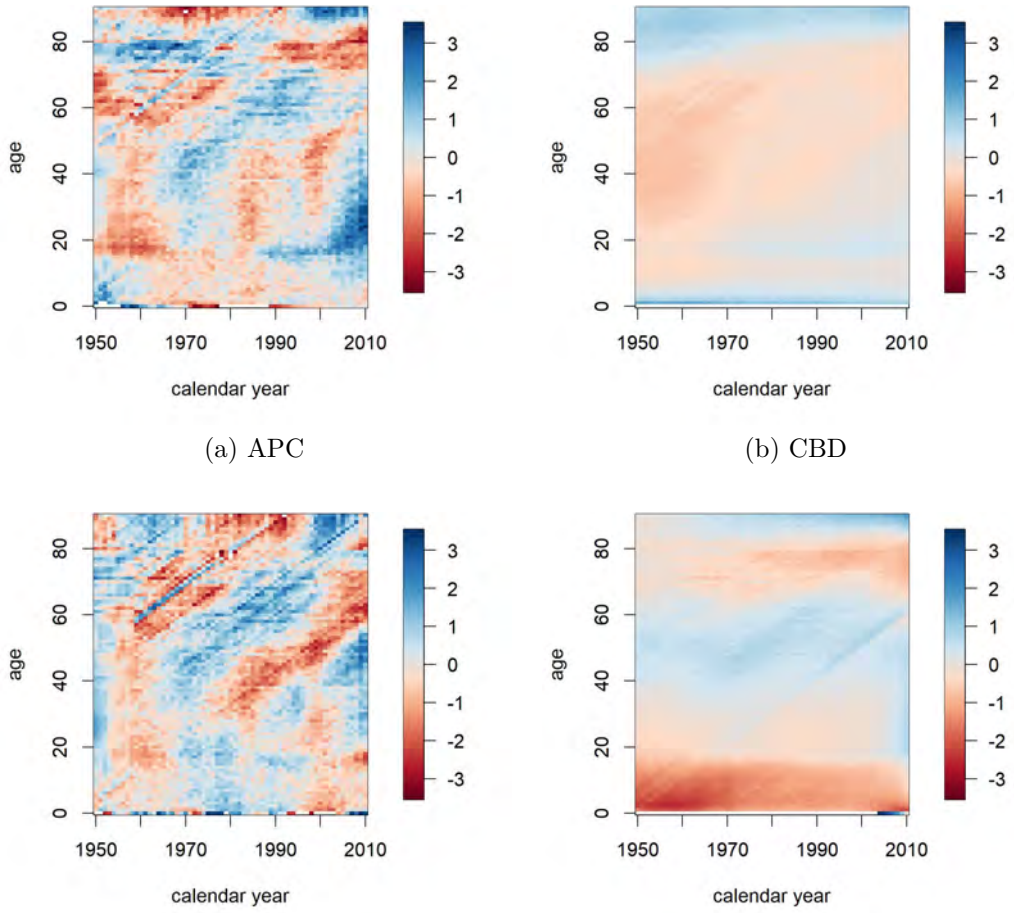


(a) Finnish females (b) German females

(c) Italian females (d) Dutch females

Figure 16: Parameters for the Age-Period-Cohort (APC) model fitted to the Finnish, German, Italian and Dutch female populations for ages 0-90 and the period 1950-2010 (1990-2010 for Germans). The period effect is projected up to year 2019.

The goodness-of-fit of mortality models is usually analyzed by examining the residuals of the fitted model. Regular clustering patterns in the residuals imply that the model is not able to adequately describe all features of the data. In the case of a Poisson setting, it is appropriate to look at the standardized deviance residuals, which were defined in Subsection 4.5. From figure 17 one can see the standardized deviance residuals for all four models presented as heatmaps for the US women. Models CBD and M7 display strong residual clustering patterns, while the residuals of models LC and APC look reasonably random, even though LC model also shows less pronounced clustering patterns indicating the inability of this model to capture the well-known cohort effect observed in the US female population.

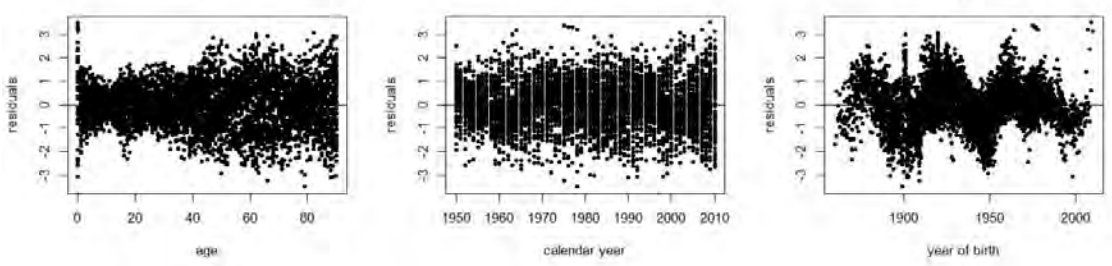


(a) APC (b) CBD

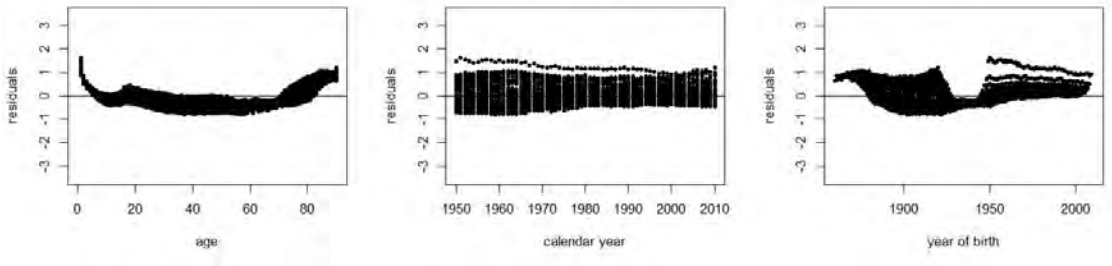
(c) LC (d) M7

Figure 17: Heatmaps of deviance residuals for different models fitted to the US female population for ages 0-90 and the period 1950-2010.

The weaknesses of the LC and CBD models in not incorporating a cohort effect become more apparent when looking at scatter plots of the residuals by age, period and cohort. The right hand panels in figure 18 highlight the inability of the LC and CBD models to capture the cohort effect. In addition, the left hand panel of figure 18 (b) shows some strong patterns by age, mirroring the lack of a quadratic age term in CBD. This may be necessary to capture the commonly observed curvature in mortality rates on a logarithmic scale.



(a) LC



(b) CBD

Figure 18: Scatterplots of deviance residuals for models LC and CBD fitted to the US female population for ages 0-90 and the period 1950-2010.

If one looks at the in-sample errors based on MAPE as a criterion in figure 19, it is apparent that the LC and APC models perform better than the CBD and M7. This phenomenon refers to all three components, age, period and cohort. It is also worth mentioning that these two models fit poorly, especially for lower age categories. Both models use the linearity assumption which holds rather for retiree age, so these models only provide reasonable fits for the higher age groups.

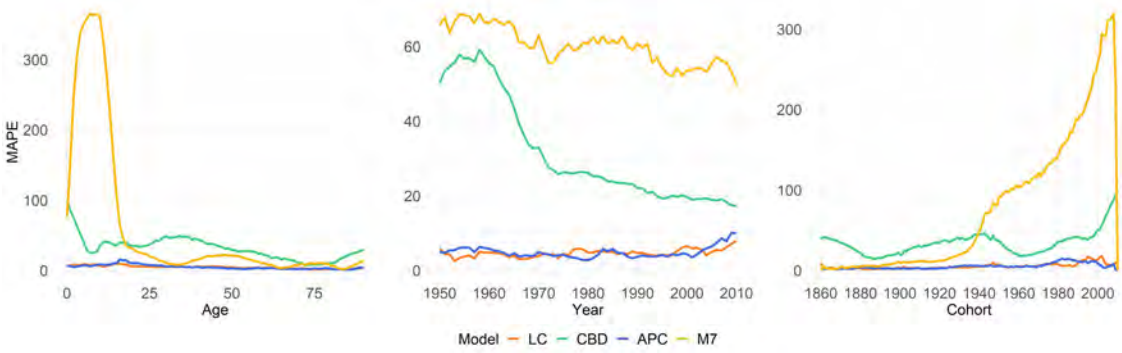
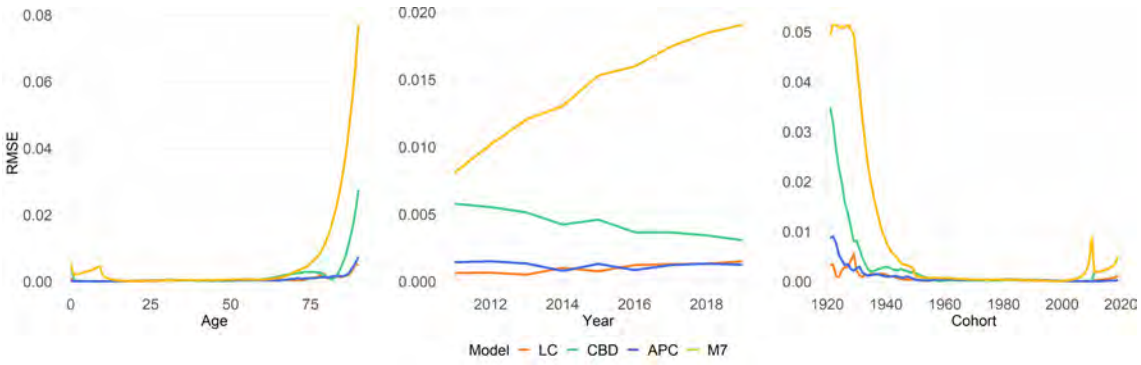


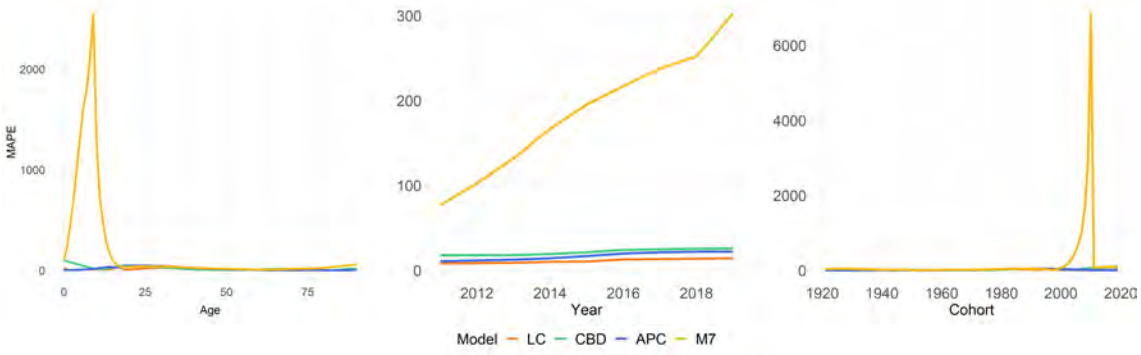
Figure 19: Lineplots of in-sample Mean Absolute Percentage Error (MAPE) for models LC, CBD, APC, M7 fitted to the US female population for ages 0-90 and the period 1950-2010, compared to the respective components Age, Period and Cohort.

When evaluating the goodness-of-fit, models with more parameters are considered to provide a better fit to the data. To exclude the possibility that the better fit observed in a model is the result of over-parametrisation and compare the relative performance of several models, one will look at the out-of-sample errors. More precisely, it means that the model is fitted on the years 1950-2010, and projected into the years 2011-2019, whereas the forecasts are then related to the actual mortality

rates for this period in an appropriate manner. Thus, the figure 20 indicates the out-of-sample errors measured on MAPE. While the disfavor for the M7 model is also clear here with respect to all three components, the poor functioning of CBD is only enforceable with respect to period. It is worth noting that the CBD model performs better on the test data, so that the errors are smaller on average than those of the training data. On the other hand, a look at RMSE in figure 20 allows us to maintain the original assumption that LC and APC models work best also in terms of forecast.



(a) RMSE



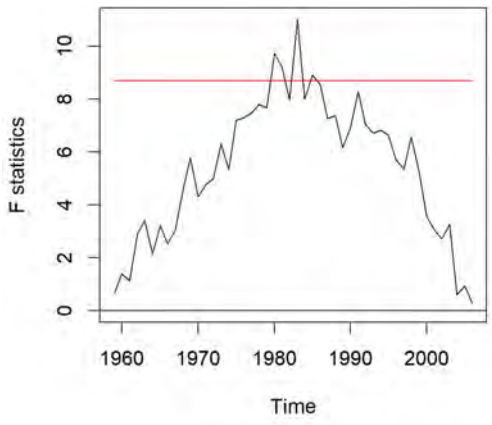
(b) MAPE

Figure 20: Lineplots of out-of-sample Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) for models LC, CBD, APC, M7 fitted to the US female population for ages 0-90 and the period 1950-2010, compared to the respective components Age, Period and Cohort. The out-of-sample forecast was conducted for the years 2011-2019.

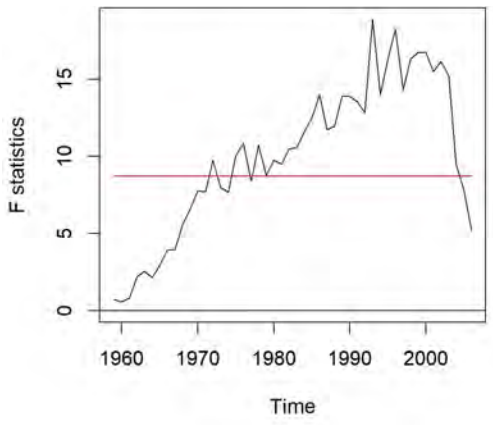
5.2 Breakpoint analysis

The period effects extracted in the mortality models are forecasted using random walk with linear drift. This assumption is not always fulfilled, so we look to see if we can find structural breaks for the linear drift. In this subsection, we explore how to use structural breaks to contribute to the stability of this method. Especially, this is important to use the models in purpose of long-term mortality and longevity forecasts. If structural changes are present, time-dependent effects, such as period effects are possibly very sensible towards the calibration period and cannot be captured by standard ARIMA-models. We will initially check for structural changes by the

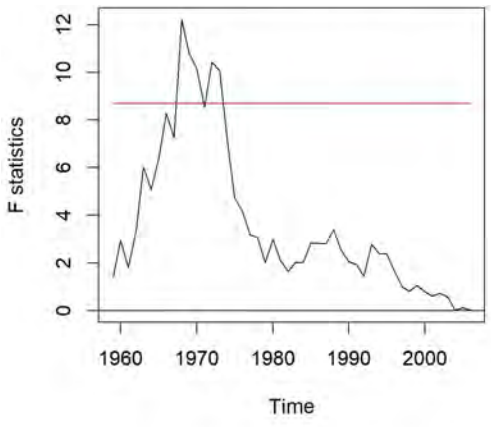
methods proposed by Bai and Perron (2003) and Zeileis et al. (2003) introduced in Subsection 4.1 for each subpopulation and its time series individually and date the breakpoints if there are present any. Given the dated breakpoints, a univariate random walk with constant drift will be then fitted and deviations identified if the period trend is changing over time. Once each model has been fitted to the data for the period 1950-2010 (in case of Germany 1990-2010) for each male and female, we first extract the corresponding time series $\kappa_{t,s}$ that presents the main period improvement factor in the Lee Carter model. Then, the first order differences are built, which provides basis for further analysis. At this point, it should be noted that for the purpose of this thesis, we use the Lee Carter model to detect structural breaks because latest research suggests that this serves as a benchmark model in this sense as well and that the main driver of mortality forecasts based on any of the above models is that of the main period effect $\kappa_{t,s}$ (van Berkum et al., 2013). Since of all subpopulations, only for the male populations in the US, the Netherlands and Italy one break each is found, the focus is further on these three. These findings can be confirmed both with the help of an F-test and using BIC by using the package `strucchange` for the implementation.



(a) ITA (males)



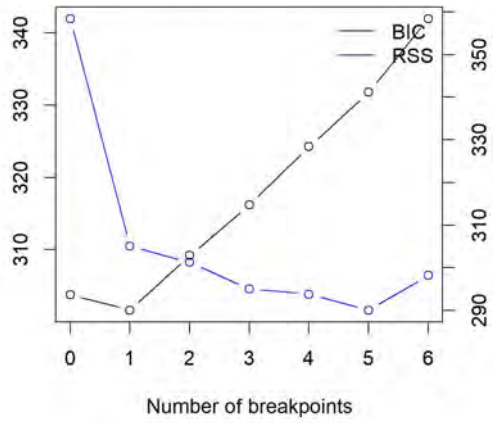
(b) NLD (males)



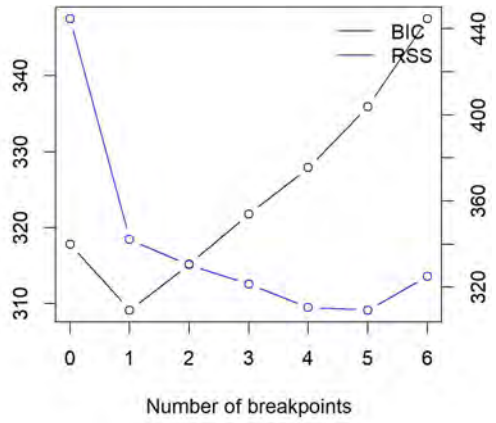
(c) USA (males)

Figure 21: F statistics crossing their boundary and thus indicating a structural change (at the level $\alpha = 0.05$).

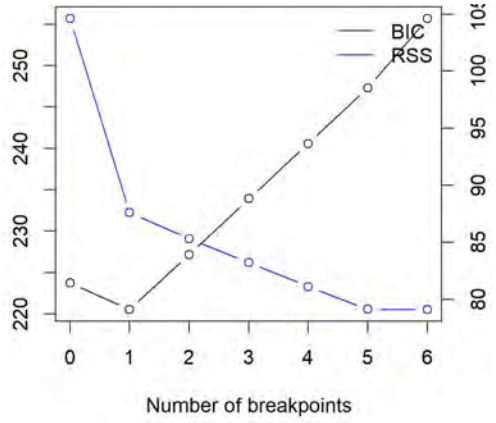
Figure 21 shows the identification of structural breaks for these three subpopulations in a certain time period each. The structural change is located and dated by plotting the F statistics for each time period with its boundary. Fluctuations that fall outside these known boundaries are judged to be improbably large and hence suggest a structural change in the mean value. It is rather similar to that of empirical fluctuation processes. The boundaries can be computed under the null hypothesis of no structural change present, such that the asymptotic probability that the supremum of the statistics F exceeds this boundary is 0.05. The alternative hypothesis is the precense of one break point. The F statistics is a function of the restricted and the unrestricted sum of squared residuals (the null and alternative hypothesis, respectively). As the F statistics cross their boundary in these three cases, there is evidence for a structural change at level 0.05. Each of them has a fairly definite pronouncement, which mirrors the results from the analysis by empirical fluctuation processes and tests, respectively. For details on calculations, please refer to Zeileis et al. (2003). On the other hand, looking at the BIC one gains a detailed information about not only the comparison no structural change versus one change is present, but on an even more granular level. We used this strategy primarily to identify the optimal number of breakpoints, which is one for all three cases, as depicted in figure 22 providing information about the estimated results for different numbers of breakpoints. As expected, the RSS decreases as the number of breakpoints increases. Since the BIC is minimal for one break point for all three cases, using the BIC as a criterion suggests using one break point only.



(a) ITA (males)



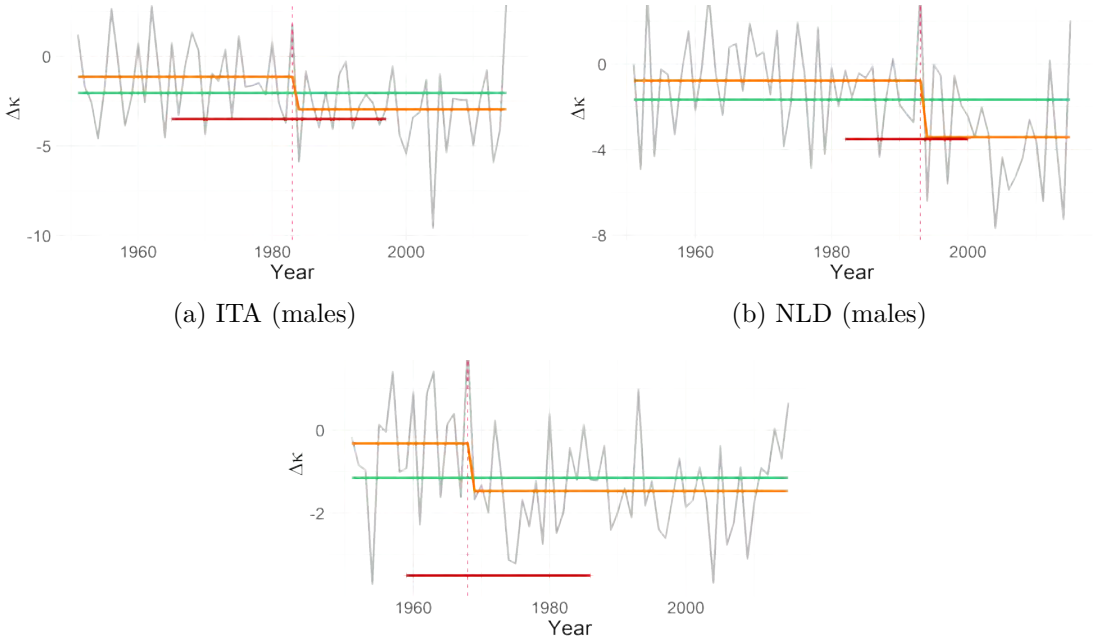
(b) NLD (males)



(c) USA (males)

Figure 22: Estimation results for different numbers (0-6) of breakpoints. The optimal value for the BIC (black lines) is in all cases 1. The blue lines indicate the Residual sum of squares.

The novel insights gained about the presence and approximate period of the breaks in the time series allows us to further deepen the recognition in the direction of concrete annual appointments for these breaks. Figure 23 shows the corresponding breakpoints as concrete values obtained with the package `strucchange`. The confidence intervals (red lines) are given indicating that the breakpoints (dotted horizontal red line) are not exact point estimates, but are within a certain range. Whereas the orange lines represent the mean of $\Delta\kappa_{t,s}$ for the different periods when one breakpoint is considered, the green line represents the mean of $\Delta\kappa_{t,s}$ when no breakpoints are allowed. For US males the break point is detected in 1968, for Italian males in 1983 and for Dutch males it is in 1993. Having more structural breaks present for males than females is in accordance with findings from the literature. Coelho and Nunes (2011) for example, found out that only 5 of the 18 examined countries have structural breaks for females suggesting that any potential acceleration in mortality improvement has had a greater impact on male mortality than on female mortality.



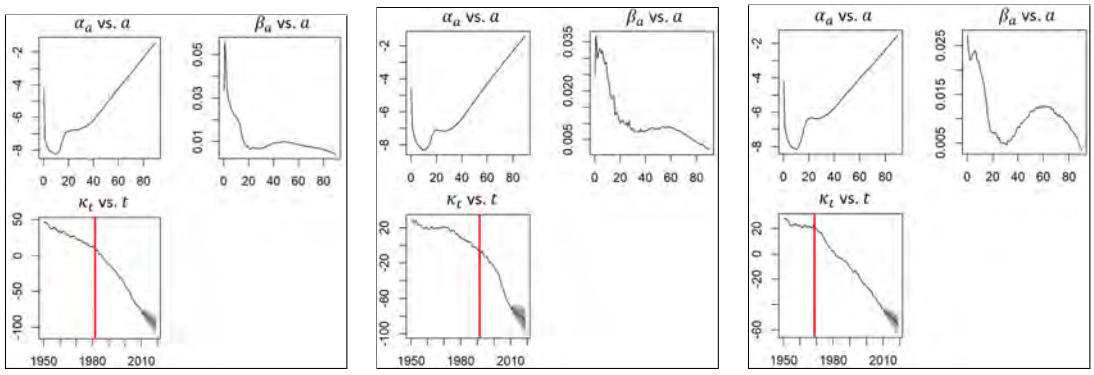
(a) ITA (males)

(b) NLD (males)

(c) USA (males)

Figure 23: Confidence intervals for estimated breakpoints for the first order difference of the period effect $\kappa_{t,s}$. In addition, the mean of the period differences is visualized with no breakpoints (green lines) and after allowing for one breakpoint each (orange lines).

The figure 24 shows that in all three cases the structural break is accompanied by improvement in mortality. That is, after these years, the period effect estimate looks even more steeper descending, indicating an accelerated improvement in male mortality. Whether the break is more or less pronounced can be reflected on the width of the confidence intervals in figure 23.



(a) ITA (males)

(b) NLD (males)

(c) USA (males)

Figure 24: Parameters for the Lee-Carter (LC) model fitted to the Italian, Dutch and US male populations for ages 0-90 and the period 1950-2010. The red lines represent the dates breakpoints.

It is difficult to trace to which events these breakpoints can be attributed. In the US, mandatory seat belts were introduced in 1968. The new model of the three-point

seat belt that most of us know today secured the chest and hips with a single belt. These seat belts became mandatory in all new vehicles in the US. The law took effect on January 1, 1968. This may have been a possible link to the break in mortality rates in the time series (Dept of Transportation (US), National Highway Traffic Safety Administration (NHTSA), 2010). In the early 1990s, a reform of the health care system was enforced in the Netherlands. This improved conditions for the health-insured and thus for the perception of many medical treatment options (Cohu et al., 2011). Studies show that there was a marked decrease in coronary heart disease mortality in Italy from 1980 onward. Changes in the prevalence of active smoking from this period onward were mainly responsible for this (Barone-Adesi et al., 2011). On the other hand, one might wonder if this can really have such a rapid effect and if it really does not affect women. However, some of our assumptions are in line with other research, claiming that factors like medical progress (Bots and Grobbee, 1996) and reforms of health systems (Moreno-Serra and Wagsta, 2010) can have an impact on the rapidity of the mortality improvements. However, establishing a precise link between concrete events and these breakpoints is not the focus of this thesis and will therefore be left at this conjecture.

Country	Female				Male							
	LC	CBD	APC	M7	LC	LC (with BP)	CBD	CBD (with BP)	APC	APC (with BP)	M7	M7 (with BP)
FIN	0.0045	0.0157	0.0044	0.0037	0.0072		0.0181		0.0060		0.0120	
DE	0.0015	0.0055	0.0006	0.0008	0.0021		0.0067		0.0011		0.0028	
ITA	0.0025	0.0195	0.0026	0.0043	0.0035	0.0012	0.0222	0.006	0.0029	0.0015	0.01	0.0026
NLD	0.0019	0.0136	0.0031	0.0030	0.0041	0.0017	0.0146	0.0048	0.0025	0.0013	0.0067	0.0031
US	0.0014	0.0123	0.0015	0.0044	0.0020	0.0017	0.0132	0.0096	0.0026	0.0018	0.0081	0.0049

Table 6: In-sample RMSE for all countries and genders and models LC, CBD, APC and M7. The fitted period is 1950-2010 for Finland, Italy, the Netherlands, the US and 1990-2010 for Germany. The reference in the brackets indicates the fitting taking into account breakpoints.

The task now is to evaluate the extent to which accounting for structural breaks in these three subpopulations leads to improvement in fit and prediction. The table 6 shows the in-sample errors measured by RMSE. At this point, it should be noted that for all tables from here on, the light blue highlighting represents the comparatively smaller column comparison errors. Apart from a clear preference for LC and APC models, it is immediately apparent that the consideration of breakpoints leads to a relevant improvement in fit resulting in better accuracy. The years 1969-2010 for US male, 1984-2010 for Italy and 1994-2010 for the Netherlands were used for training and evaluation. This is mainly due to the fact that a more recent trend will better reflect future mortality rates and it is noted that the results are varied with very minor improvement in some cases.

Country	Female				Male							
	LC	CBD	APC	M7	LC	LC (with BP)	CBD	CBD (with BP)	APC	APC (with BP)	M7	M7 (with BP)
FIN	0.0020	0.0051	0.016	0.0058	0.0029		0.0067		0.021		0.0063	
DE	0.0048	0.0077	0.011	0.0065	0.0052		0.0094		0.016		0.0070	
ITA	0.0045	0.0052	0.019	0.0071	0.0045	0.0042	0.0048	0.0075	0.026	0.015	0.0074	0.0076
NLD	0.0030	0.0063	0.018	0.0054	0.0047	0.0035	0.0045	0.0082	0.028	0.018	0.0066	0.0065
US	0.0023	0.0052	0.016	0.0029	0.0052	0.0049	0.0076	0.0084	0.021	0.017	0.0049	0.0048

Table 7: Out-of-sample RMSE for all countries and genders and models LC, CBD, APC and M7. The fitted period is 1950-2010 for Finland, Italy, the Netherlands, the US and 1990-2010 for Germany. The forecast period is 2011-2019. The reference in the brackets indicates the fitting taking into account breakpoints.

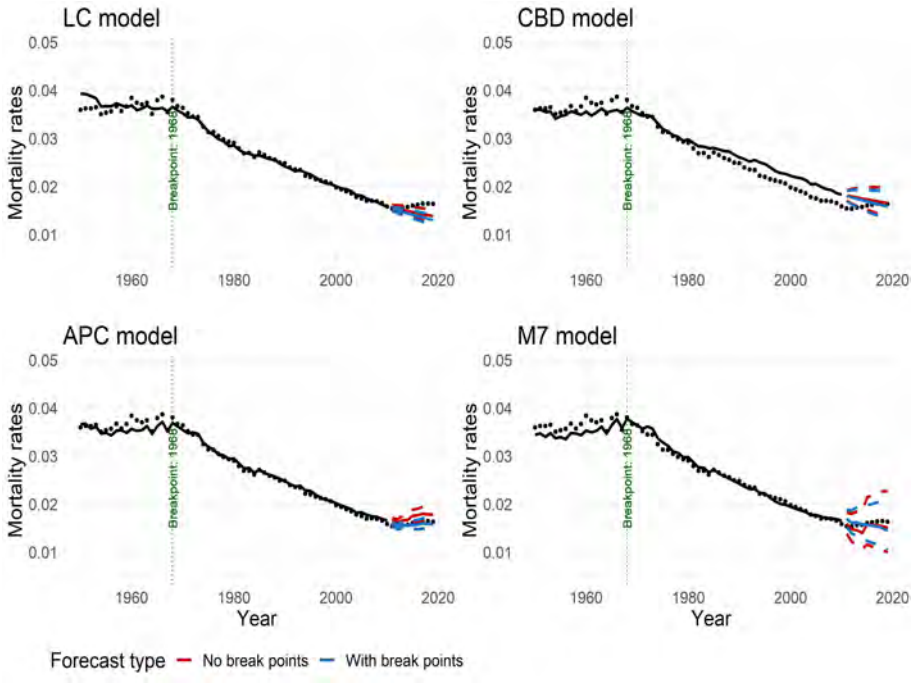


Figure 25: Mortality rates fitted and forecasted for 65-year-old US male population, calibrated on the years 1950-2010. The red lines correspond to forecasts when breakpoints are not allowed for, the blue lines correspond to forecasts when breakpoints are allowed for. The dashed lines stand for the 95-%-confidence intervals for the forecast for years 2011-2019. The black points indicate the observed mortality rates.

It is even more important to investigate the effect on the out-of-sample errors. This is to test how well the predictions work on the years 2010-2019, which are unseen by the model. Again, the improvement is not negligible with respect to all four models and three subpopulations. This is demonstrated in the table 7, indicating that mortality forecasts using the proposed method are more robust towards the calibration period. A more detailed exploration of both fit and forecast of all four models with and without breakpoints in visual form are allowed by the figures 25 and 26 These show the actual and predicted mortality rates for 65 and 85 year old US men as an example.

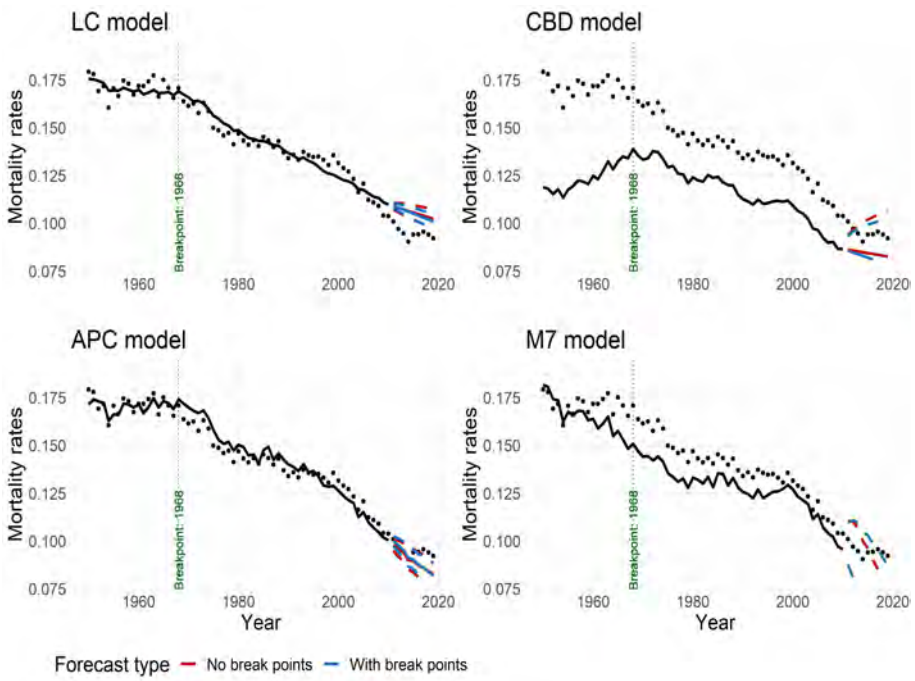


Figure 26: Mortality rates fitted and forecasted for 85-year-old US male population, calibrated on the years 1950-2010. The red lines correspond to forecasts when breakpoints are not allowed for, the blue lines correspond to forecasts when breakpoints are allowed for. The dashed lines stand for the 95-%-confidence intervals for the forecast for years 2011-2019. The black points indicate the observed mortality rates.

It is confirmed that the LC and APC models fit and forecast better compared to CBD and M7. This is visible from the degree of fit from the line at the black dots representing the actual observations. LC supports linear extrapolations and incorporating historical beakpoints does not help too much. APC is much more flexible, local adaptive and thus better encompasses the actual observations in the future. The dashed lines represent the 95%-confidence intervals, which are relatively wide and do not provide large changes depending on whether modeling is done with or without a breakpoint. With the APC model, however, the improvement after considering beakpoints is visible.

5.3 Machine learning methods

From the last two subsections, the following two key messages were concluded: first, it became apparent that the LC and APC models perform best for almost all countries and genders, while M7 and CBD models tend to provide good model and predictive performance only for older age categories. In addition, taking into account structural breaks and, if such breaks exist, adjusting the years to be trained accordingly will certainly improve the fit and forecast in the given cases. Next, we revisit the LC model under consideration of structural breaks to further improve the accuracy and to analyse its calibration to the morality data from all of the subpopulations covered in this thesis. So, in this subsection, we aim to demonstrate how machine learning techniques allow us to further examine the adequacy of the estimated mortality rates by the LC model. The ideas presented in Subsection 4.2 are now applied according to Deprez et al. (2017) and Levantesi and Pizzorusso (2019), by using

tree-based approaches to analyze how modeling should be improved based on feature components of an observation, such as age or cohort.

The implementation starts with the Lee Carter model taking into account detected breakpoints, when available, which is fitted to the data with the help of `StMoMo` package for a given country and gender. The aim is then to evaluate the estimated mortality rates $\mu^{LC}_{a,t,s}$. To do this, three types of supervised tree-based machine learning models are used to calibrate the machine learning estimated improvement factor $q_{a,t,s}$ using age, year and cohort as input features: Decision Tree (DT), Random Forest (RF), and Gradient Boosting Machine (GBM). In stage one, we rely on the classical form of the LC model and improve it in the second step by using machine learning and by multiplicatively introducing the improvement factor. The idea is rather complementary and not alternative to the standard mortality models. The relationship established in Subsection 4.2 to obtain the improvement factor is:

$$\begin{aligned} D_{a,t,s} &\sim Poisson(d_{a,t,s} \cdot q_{a,t,s}), \\ &with\ d_{a,t,s} = E_{a,t,s} \cdot \mu^{LC}_{a,t,s} \end{aligned} \tag{52}$$

In stage two, we make use of the machine learning framework to find $q_{a,t,s}$ as a solution of the tree-based algorithms applied to the ratio between the death observations and the death counts estimated by the LC mortality model as suggested by Levantesi and Pizzorusso (2019):

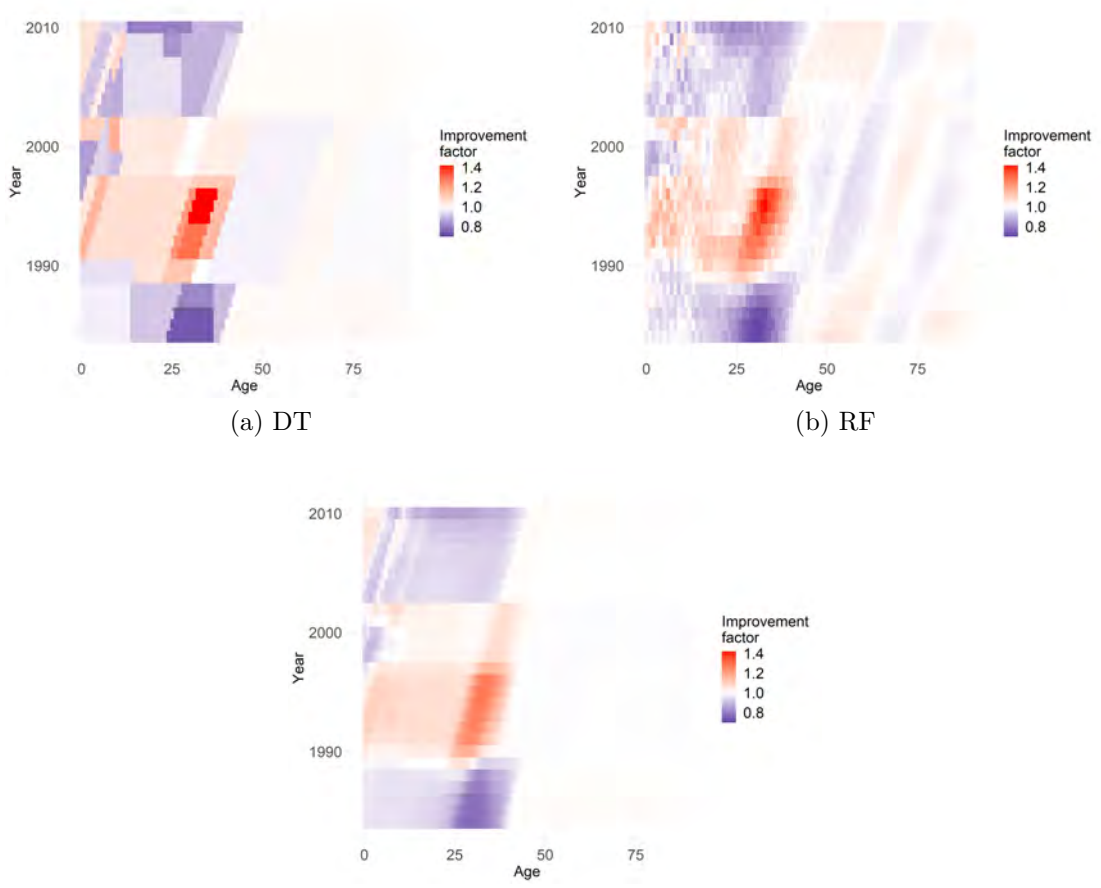
$$\frac{D_{a,t,s}}{d_{a,t,s}} \sim age + year + cohort \tag{53}$$

The introduction of the cohort effect may pose identifiability issues, but the authors justify this by saying that they do not aim to explain or interpret the effects, but rather to make appropriate predictions. In order to implement the machine learning algorithms explained in Subsection 4.2 the packages `rpart`, `randomForest` and `gbm` were used. The settings of the important parameters are based on recommendations given by Levantesi and Pizzorusso (2019). For Decision trees, we set the complexity parameter to $cp = 0.003$, for Random Forest we put the number of trees to $ntrees = 200$, to ensure that an appropriate proportion of the variance is explained while producing a low mean of the squared residuals. For gradient boosting we set the number of trees to $n.trees = 5000$, number of cross-validation folds $cv.folds = 5$ and learning rate $shrinkage = 0.001$.

Figure 27 shows the values of the fitted improvement factors $q_{a,t,s}$ by age and year for Italian male population. White areas represent value 1 for the improvement factors indicating that the specified mortality model, in this case the LC model perfectly fits the mortality rates, introducing only very small variations from the Lee Carter model. Values smaller than 1 (purple) point out that the mortality model could overestimate and the values higher than 1 (dark orange) could underestimate the actual mortality rates. It is apparent that many areas are identified by diagonal splits, highlighting the cohort effect, justifying the decision to insert the cohort parameter additionally in the tree-based algorithms.

Country	Female				Male			
	LC	Tree	RF	GBM	LC	Tree	RF	GBM
FIN	0.0045	0.0021	0.0027	0.0011	0.0072	0.0019	0.0033	0.001
DE	0.0015	0.0014	0.0006	0.0004	0.0021	0.0019	0.0008	0.0006
ITA	0.0025	0.0007	0.0019	0.0008	0.0012	0.0006	0.0008	0.0003
NLD	0.0019	0.0004	0.0011	0.0002	0.0017	0.0000	0.0008	0.0000
US	0.0014	0.0007	0.0012	0.0005	0.0017	0.0012	0.0014	0.0006

Table 8: In-sample RMSE for all countries, genders and models LC as well as the tree-based improvements with Decision Tree, Random Forest and Gradient Boosting Machine Models. The fitted period is 1950-2010 for Finland, Italy, the Netherlands, the US and 1990-2010 for Germany.



(a) DT

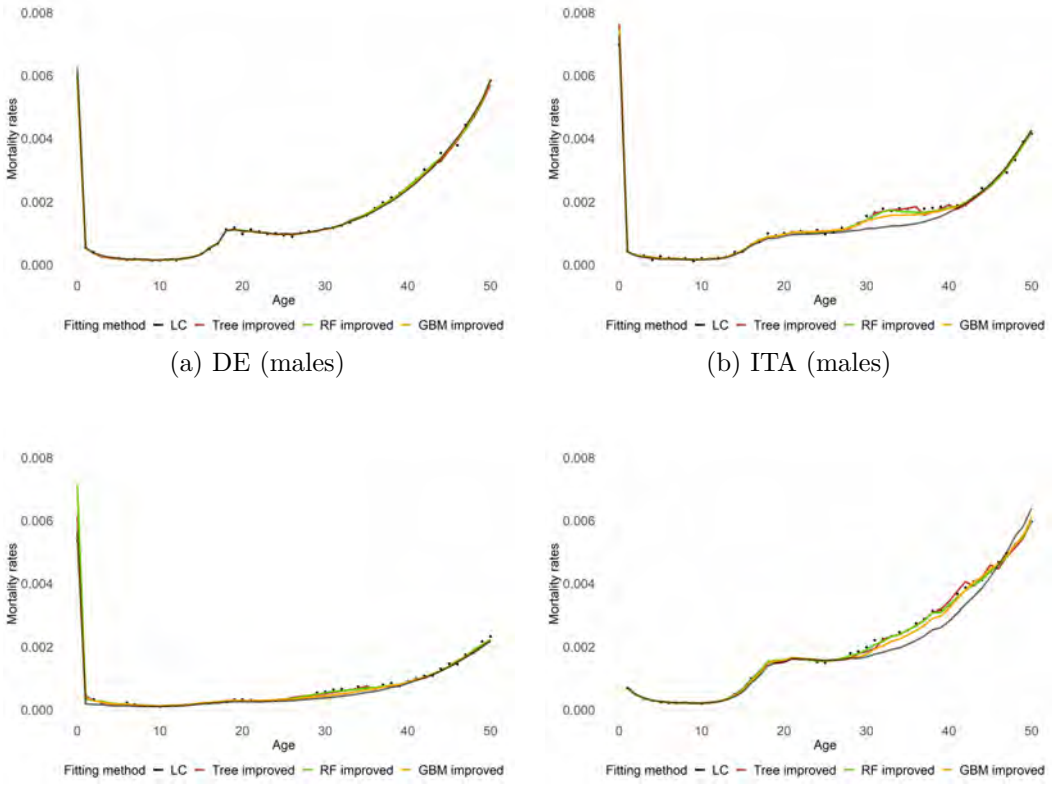
(b) RF

(c) GBM

Figure 27: Values as heatmaps of the improvement factors for each year (1950-2010) and age (0-90) achieved with Decision Tree (DT), Random Forest (RF) and Gradient Boosting Machine (GBM) model for Italian male population..

Figures 28 represent the actual and fitted mortality rates only for younger ages, since the most relevant changes seem to be concentrated in the younger ages, as seen in the previous figure. This holds for other subpopulations as well. For selected subpopulations, we see in figures 28 first the actual mortality rates as black dots exemplary for the year 1995. The black line shows the fit of the LC model based

on the years up to and including 2010. The red, green and yellow lines, on the other hand, show the fit improved by Decision Tree , Random Forest and Gradient Boosting, respectively. It is obvious that the fit to the actual points or mortality rates improves by multiplying by the respective improvement factors. While the black line is further away from the actual values in some places, often for the even younger age categories, the improved methods are closer. A deeper analysis of goodness-of-fit is provided in the table 8. It presents the subpopulations row by row and the classical Lee Carter model as well as the improved versions column by column. A look at the RMSE makes clear that an immense reduction of errors could be achieved by all machine learning methods compared to the original LC model, especially for the boosting method. In almost all cases Gradient Boosting Machine (GBM) achieves the highest reduction of RMSE and thus better fit, followed by Decision Trees (DT) and Random Forests (RF).



(a) DE (males)

(b) ITA (males)

(c) ITA (females)

(d) USA (males)

Figure 28: Mortality rate forecasts for selected subpopulations, 1950 and ages 0-50. The black points indicate the observed mortality rates. The black lines represent the classic LC model estimations, whereas the red (DT), green (RF) and yellow (GBM) lines improved the LC estimates with the corresponding machine learning method.

From the graphs, one can argue that machine learning estimators led to an improvement in the quality of fit in the mortality model. Thus, it can be concluded that the application of machine learning methods improves the mortality rate estimations locally.
Furthermore, it is important to test not only the goodness-of-fit, but also the ability of the improved models to be generalized. For that the mortality rates estimations

improved by the machine learning estimator are forecasted according to the Lee Carter framework, allowing one to assess the forecasting quality.

Country	Female				Male			
	LC	Tree	RF	GBM	LC	Tree	RF	GBM
FIN	0.0021	0.0021	0.0014	0.0017	0.0029	0.0034	0.0024	0.0031
DE	0.0048	0.0046	0.0032	0.0033	0.0052	0.0052	0.0033	0.0034
ITA	0.0045	0.0046	0.0039	0.0022	0.0042	0.0041	0.0020	0.0022
NLD	0.003	0.003	0.0015	0.0017	0.0035	0.0035	0.0018	0.0020
US	0.0023	0.0023	0.0010	0.0011	0.0054	0.0052	0.0049	0.0051

Table 9: Out-of-sample RMSE for all countries, genders and models LC as well as the tree-based improvements with Decision Tree, Random Forest and Gradient Boosting Machine models. The fitted period is 1950-2010 for Finland, Italy, the Netherlands, the US and 1990-2010 for Germany. The out-of-sample years are 2011-2019.

The years 2011-2019 are used for testing. Again, we deliberately omit 2020 because, as discussed in detail in the section on trend forecast, this year represents a kind of outlier at the edge of a time series. The table 9 provides an overview of the out-of-sample errors and thus of the ability to generalize by relating actual mortality data to predictions that are reached based on unseen data not used to train the model. The reduction in RMSE in this case compared to the LC model is not as large as in the in-sample expansion case. It becomes clear that the Lee Carter Model is well suited to depicting temporally homogeneous time structures and to foreseeing the linear overall trend. The trees, on the other hand, manage to resolve locality better. The transformation in forecast of the local improvement is not always target-oriented. This finding may be related to an overfitting issue that occurs when a model learns the details and noise in the training data to an extent that it negatively affects the model's performance on new data. One needs to build the models in such a way that they generalize well to the new data meaning that the model needs to have the ability to properly fit to new, previously unseen data that comes from the same distribution as the one used to build the model. By achieving a too high fit to the trained data, the generalizability decreases. This unbalances the trade-off between bias and variance. This challenge was also addressed in the introduction of these machine learning algorithms and discussed comparatively. To summarize, Decision trees can be an effective statistical tool. However, they have some drawbacks. Most critically, they can overfit the data, so that a single tree poorly predicts future data that was not used to build the original tree. This problem is known in statistics as the bias-variance dilemma. Random Forest can help to overcome the overfitting problem. At the top level, Random Forest is a collection of decision trees, built slightly differently on the same dataset, that collectively tune to produce a better model than a single tree can. The trees are built by randomly selecting a subset of visit records with substitutions (known as bagging) and by randomly selecting a subset of the attributes, so that the forest consists of slightly different decision trees. This method allows for small variations in the trees that are created in the Random Forest. Adding this controlled variance helps to improve the predictive accuracy of the algorithm (Hastie et al., 2016). This can possibly provide an explanation as to why the largest decrease in RMSE is achieved in the Random Forests model.

In summary, in the case of applying machine learning techniques to our data, we were able to achieve an immense improvement in fit and a comparative improvement in forecast compared to the original estimated results of the mortality rates of the LC model.

5.4 Generalized additive models

The next class of models is the generalized additive model, which was methodologically introduced in Subsection 4.3. The concrete application refers to one of the two better performing classical mortality models examined in Subsection 5.1, namely the APC model. Here, the basis for estimating mortality rates is the combination of features age, period and cohort, assuming that these play an important role when explaining human mortality. We aim to apply another statistical approach that will allow for multipopulational modelling to project and predict future mortality improvements in a coherent way, without contraints. Hereby, we assume that jointly learning from each other both between countries and between genders within each country will contribute to the model performance. For this we utilize GAMs as extensions of GLMs resulting in a more robust and flexible method for capturing non-linear effects. The identifiability problem arised by three components of the temporal structures age, period and cohort representing linear combinations of each other is tackled using a non-linear interaction surface with a two-dimensional tensor product basis, capturing the cohort effect implicitly. To implement the models of this subsection the R package `mgcv` by Wood (2021) was used. The number of knots were set to ten in an equidistant way in both dimensions. The fitted years are in this section 1990-2015, as we make use of the joint years present in the dataset, which is limited to 1990 due to Germany mainly. The range used for testing comprises the years 2016-2019. The specialities concerning the year 2020 will be analysed in the next section. In the first step, we describe the basic constellation as a reference for the evaluation to see if there are pronounced country and gender effects in the more complex model. We refer to this model equation

$$\log(\mu_{apc,i}) = \beta_0 + f_{ap}(age_i, period_i) + \log(E_i) \tag{54}$$

and do not model countries and genders differently, assuming that there are no relevant disparities. However, we can look at the marginal effects of age, period and cohort from the general point of view (figure 29).

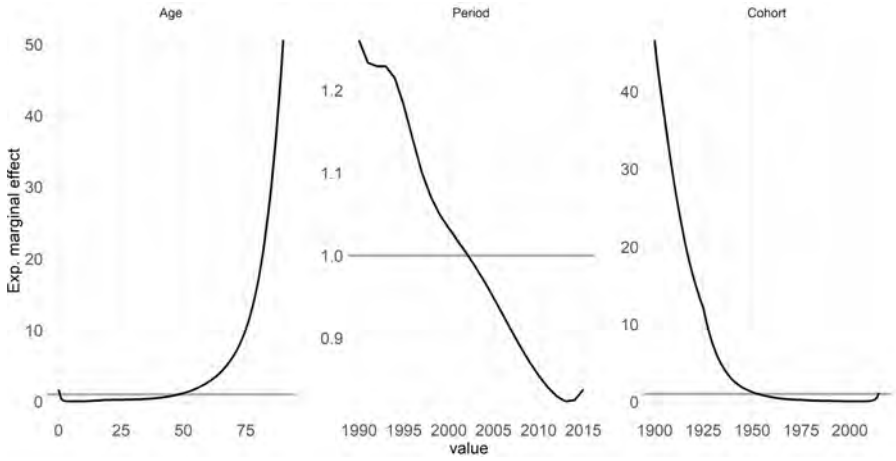


Figure 29: Estimated marginal effects of age, period and cohort for mortality rates. The horizontal lines mark the level of having no effect at all. The GAM was fitted for years 1990-2015, for ages 0-90.

The multiplicative effect of age on mortality rates decreases from birth to 1 year old and remains negative (below 1). For people older than 50 years, this effects starts to increase reaching the pick of acceleration for people over 75 years of age. This is in line with the descriptive analysis indicating the infants having higher mortality rates compared to more adult ages. Even older people who are closer to retirement age naturally have even higher mortality rates. The period effect shows a decreasing trend almost throughout. This decrease is in line with observations from recent centuries that mortality has declined in all age groups, leading to an increase in life expectancy. This decline is due, among other things, to the decrease in mortality from infectious diseases as well as from cardiovascular diseases, cancer and overall medical progress. With two exceptions, one at the beginning of the 90s, there seems to be a small constant part suggesting that there is no period effect here. Furthermore, from the year 2014 onwards, there is a slight increase in the period effect, which probably indicates that mortality has increased in recent years compared to the years before. It will be analyzed and discussed later. It is worth mentioning that this is a comparative mean effect that applies, given everything else, especially age remains the same when comparing. The cohort effect is a kind of ex-post stratification, because the cohort was only implicitly considered in the modeling itself. A clear decline in the effect on mortality rates can be seen from the 1950 birth cohort onward. Here, too, a small increase can be seen for the last generations, which, however, hardly reaches the range of the positive effect.
All these effects have to be interpreted as a multiplicative factor on expected average mortality rates, given all other features remain constant. Since the exposure of the population is also assumed to be constant, these factors can be interpreted as effects on the death counts in same fashion.
Now comes the move to the more complex model. This has been confirmed in the descriptive section and will be seen also later. Having gained an initial glimpse here of the directions and trends of the individual components, it is beneficial to study the effects at a more granular level. Therefore we now allow for individual smooth function estimation by the interactions of the tensor product with each subpopulation yielding in the following model formulation, which now contains the

individual subpopulations s_i additionally compared to the model from equation 54:

$$\log(\mu_{apc,i}) = \beta_0 + f_{ap,s_i}(age_i, period_i) + \log(E_i) \tag{55}$$

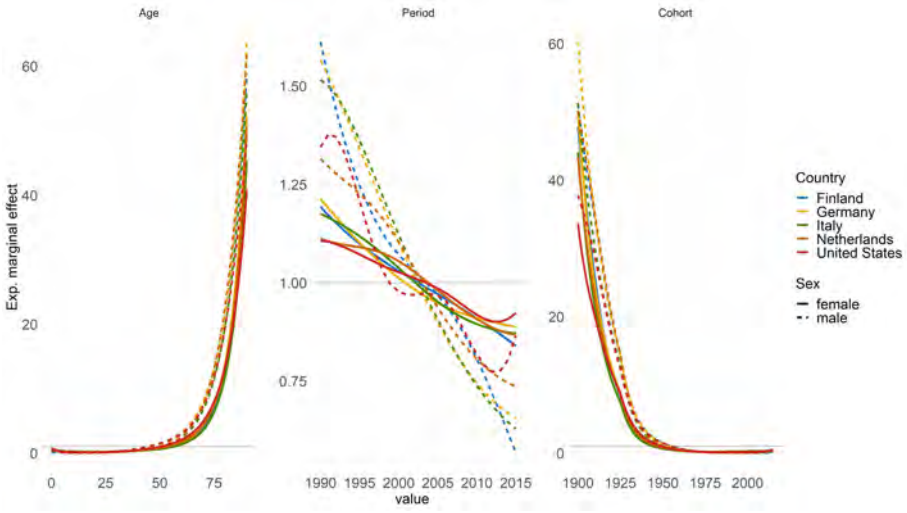


Figure 30: Estimated marginal effects of age, period and cohort for mortality rates, considering the countries and genders in the model. The horizontal lines mark the level of having no effect at all. The GAM was fitted for years 1990-2015, for ages 0-90.

The figure 30 illustrates the marginal effects for age, period, and cohort, this time differentiated by country and gender. The explicit modeling of age and period and the subsequent extraction of cohort effects allows to disentangle the effects attributed to the individual components here as well.

Age:

Age effects can be viewed as variations linked to biological and social processes of aging that are specific to individuals. They include physiological changes and the accumulation of social experiences that are associated with aging but are not related to the time period or birth cohort to which an individual belongs. In epidemiological studies, age effects are usually characterized by different rates of disease or mortality rates in different age groups (Reither et al., 2009). The age effects show overall the same structure as described for the pure model according to figure 29. The strongly regulated but complex age-specific pattern of all-cause mortality has long suggested the possibility of a universal law of mortality (Carnes et al., 1996, Olshansky and Carnes, 1997). The most important finding in this context dates back a very long time. This is arguably Gompertz's law, which shows an exponential increase in mortality with age (Gompertz, 1825). At the population level, it is possible to make hypotheses like health changes in old age are initiated by physiological dysregulations such as increased blood pressure and rising concentrations of total cholesterol. These changes are followed by an increase in diagnoses of diseases and conditions, both physical and mental, which are then followed by an increase in disability and

loss of physical function, and eventually death. For the individual, however, the sequence of the process may be different, as he or she does not experience some of the dimensions, and reversals of the process may also occur (Crimmins et al., 2019). There are pronounced differences in levels between males and females especially for ages between 50 and 90. This refers to the males having a higher effect for these age categories compared to females. With respect to the different countries, no relevant variations in age effects can be identified. Qualitatively speaking, this means that in all countries, men aged 75, for example, are expected to have higher mortality rate improvements compared to 25-year-olds than women in the same age comparison, conditional on the same year. Almost all modern data available on this subject show that females have a longer life expectancy than males, and the exceptions to this rule result from unusually harsh living conditions for women. This female superiority in survival is very much in evidence in old age, where females outlive males by an ever-increasing ratio as they age. Apart from that plenty of evidence in other studies clearly supports the view that in societies where there is no serious discrimination against one sex or the other, women tend to live longer than men. Since this must be a genetic factor, a result of evolution, it is almost certainly present in all human populations (Kannisto, 1996).

Period:

Generally, period effects do not result from individual but rather from external factors that affect all age groups equally at a given point in time. They can be caused by a variety of environmental, social, and economic factors, e.g., war, economic crisis. Methodological changes in outcome definitions, classifications, or the method of data collection can also generate period effects in the data (Keyes and Li, 2019). In contrast to the age effect, the period effect shows drastic differences not only between genders but also between countries. First, it can be observed for all subpopulations that the effect structure on mortality rates is descending and reaches the multiplicative factor 1 around the year 2004, indicating that around this year the effect on mortality rates should have been quite small. While the descending trend for women is quite similar and comparatively shallow between all countries, a much steeper descent is apparent for men, in addition to country differences, with the initial level even higher than for women. In other words, the differences in mortality rates between the years 1995 and 2010, for example, are even greater for men than for women, age being equal. This is an indication that the improvement in mortality rates over the years may have been stronger and more significant for men. We have noticed a similar aspect before in the course of the breakpoint analysis, where we found out and supported with literature that probably certain events over the years caused a stronger improvement of mortality rates for men. This finding, in turn, may be congruent with the observation in the data description that the gap between male and female mortality rates has narrowed over time, particularly in the older age categories. More recently, the decline in deaths from adult heart disease in many countries has contributed significantly to the overall decline in premature mortality. Thus, the reduction in mortality can be attributed to two more recent factors: more frequent overcoming of cardiovascular disease in the elderly and prevention of deaths from low birth weight in infants. The prevalence of smoking can also be attributed to causes of mortality, although the prevalence in men has declined in recent years.

Cause-specific mortality rates have also declined in both sexes, although to different degrees. Absolute declines in deaths caused by diseases of the circulatory system have been greater for men, and cancer mortality declined faster for men than for women. Injury-related deaths also declined more for men than for women. Of course, these are all possible reasons and assumptions that could cause the faster decline in the period effect in men (Rosella et al., 2016). Furthermore, the gender gap in mortality rates has narrowed in the last half of the past century, as male mortality rates have improved more rapidly than female mortality rates. Thus, the available data suggest that there are not only biological but also behavioral differences between the genders, and that social and psychological factors play an important role in differentiating the mortality patterns of women and men. For this reason, many analyses make the simplifying assumption that women and men are two distinct, independent populations (Perls and Fretts, 1998).
A closer look at the estimated effect curves of the US, in particular those of men, reveals that starting in approximately 2014-2015, an increase in the effects can be seen, compared to other countries. It is also likely that the increase in the pure effect can be attributed to these subpopulations which are shown on averaged basis in the figure 29. It is possible that this phenomenon is related to the victims of the opioid crisis, which caused the sharp increase in the number of drug addicts and deaths related to the abuse of opioid painkillers in the US (Dowell et al., 2017).

Cohort:

Cohort effects are variations that result from the unique experience of a group of individuals (cohort) based on year of birth as they move over time. Here, we are talking about risk for a health outcome described based on year of birth. Thus, a cohort effect occurs when the distribution of diseases resulting from an exposure affects age groups differently (Keyes and Li, 2019). It is here conceptualized as an interaction due to a period effect that is experienced differently by an age-specific susceptibility to a potential event or that cause. The cohort effect, similar to age, shows differences worth mentioning only with respect to gender. The generations born between the years around 1910 and 1940 show higher cohort effects on mortality rates for men than for women. For all countries and genders, however, the effect decreases for the more recent birth cohorts and remains almost constant. It is also clear that for males and females aged 60 years or older, the highest mortality rates correspond to the youngest cohorts of birth (central panels). That is, for the oldest age groups, rates increase with the cohort of birth.
All in all, the differences between countries and the changes over time make it clear that the gap between male and female mortality rates depends to a large extent on people's circumstances and epidemiological conditions related to mortality. This could include for example the dominance of diseases, public health infrastructure and health care resources. During the period when parity shifted between male and female mortality rates at older ages, and male mortality rates were twice as high, chronic diseases-particularly cardiovascular disease and cancer-have replaced infectious diseases as the leading cause of death. However, the change in the relative levels of male and female mortality rates reflects not only epidemiological changes in the distribution of causes of death over time, but also differential changes in behavioral patterns and risk exposure between men and women (Crimmins et al.,

2019). The graphs representing the estimated mortality rates for all subpopulations are provided in figure 31. The model outcomes are summarized by the heatmaps corresponding to the surface obtained from the tensor products. To ensure an overall compact view, rates were averaged in 5 year steps for age and period. The corresponding heatmaps reflect the distributions of estimated mortality rates across years and age categories for each gender (columns) and country (row).

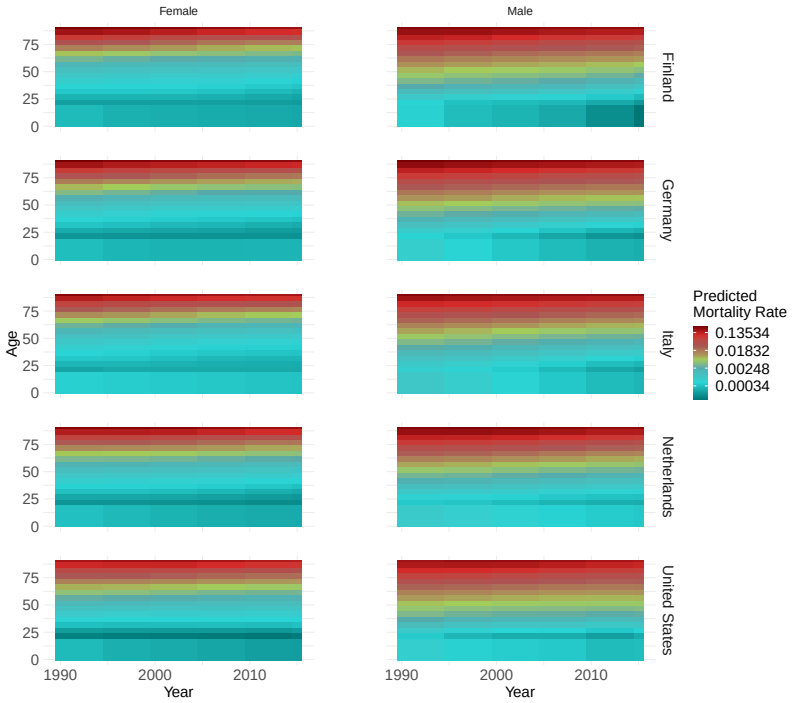
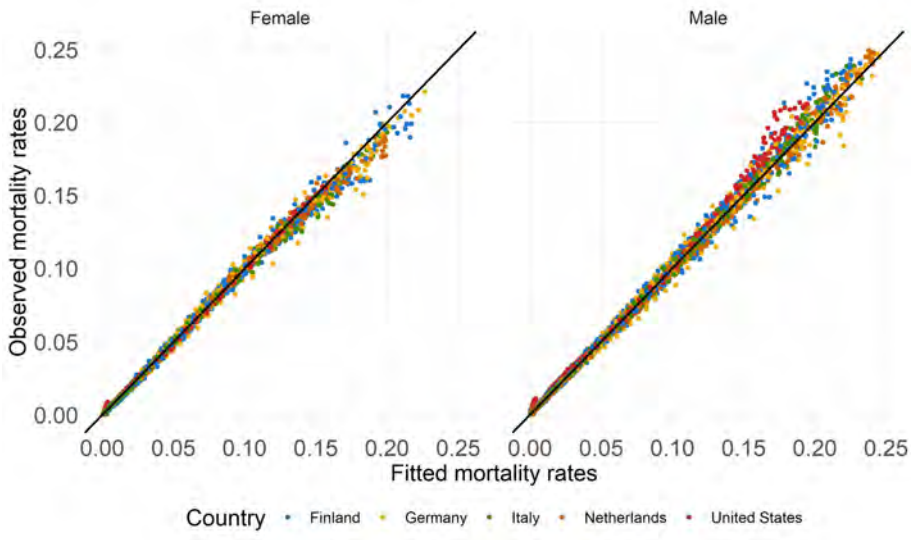
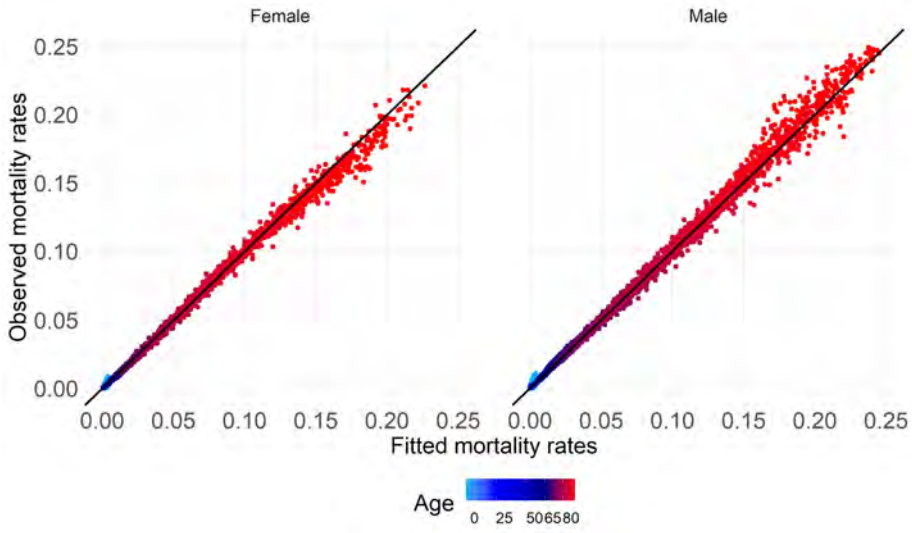


Figure 31: Heatmaps of the estimated mortality rate surface (left column: Females, right column: Males). The predictions are averaged over 5-year and -age blocks.

The figure broadly mirrors the proportions and distributions already observed in the descriptive analysis. As expected, low mortality rates are predicted for lower age categories, in particularly declining in recent years. As age increases, estimated mortality rates also increase. Moreover, women are expected to have lower rates than men. This already gives an indication of a proper model fit.



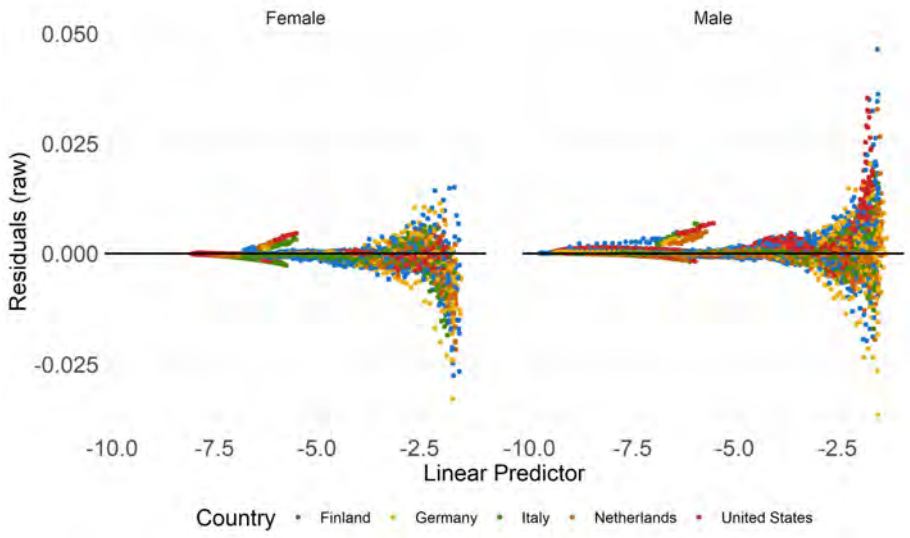
(a) Country-stratified



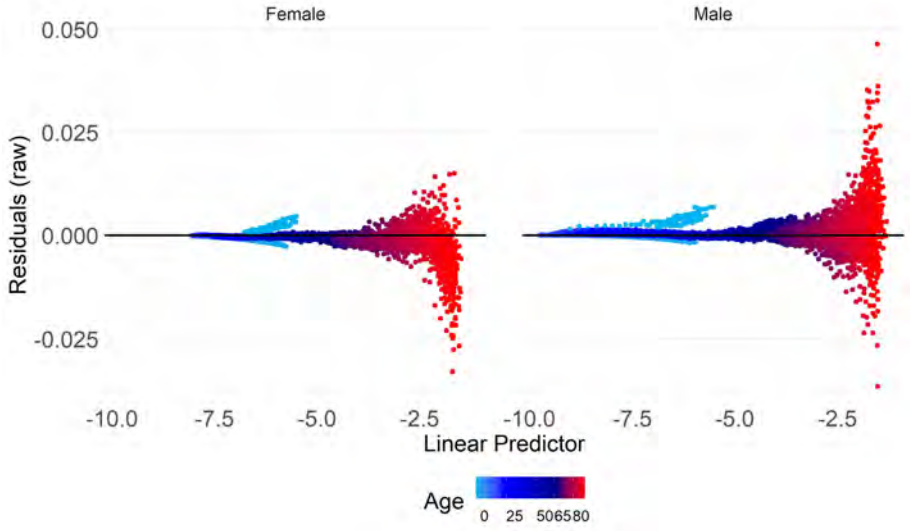
(b) Age-stratified

Figure 32: Fitted vs. observed mortality rates stratifited by country (a) and age (b).

A more in-depth analysis of model diagnostics is presented in the figures 32 - 33. On the one hand, these contrast the fitted values with the actual observed values for both males and females, at the level of crude mortality rates. Ideally, the points of the scatterplot should lie on the 45° line, which has been marked by a black line. The plots show a mostly satisfactory fit, except for slightly stronger variabilities for higher mortality rates. If this graph is stratified by country in addition to gender, regularities in the form of patterns are recognisable in the case of US. Here, the US men are almost always above the black 45° line, which indicates that the model systematically overestimates the mortality rates in this case (figure 32 (a)). However, when looking at the differences by age (figure 32 (b)), it is noticeable that for very low age categories a deviation from observed values for both sexes can be seen, these are the light blue dots. Moreover, higher variability is present in older age categories.



(a) Country-stratified



(b) Age-stratified

Figure 33: Residuals (raw) vs. linear predictor stratifited by country (a) and age (b).

If we now plot the raw residuals against the linear predictor, we again see the overestimation of US male mortality rates that is evident as they are almost systematically above the zero line (figure 33 (a)). However, the age effect seems to be relevant here as well. Deviations from zero are again caused by very young age categories which are of primary interest. Thus, the results indicate that, apart from the small deviations from the variance homogeneity assumption, the model quality is largely satisfactory. The increasing variance at higher mortality rates may be due, on the one hand, to the fact that the assumption underlying the Poisson distribution that mean and variance are equal is violated. On the other hand, it may also be because less data is available for estimating the higher age groups.
Apart from describing the relationship between the expected number of deaths and predictors such as age, period (year of death), cohort (year of birth), other lifestyle factors such as obesity prevalence, unemployment rate, tax revenue and alcohol consumption are included as linear effects, yielding in the following model:

$$\log(\mu_{apc,i}) = \beta_0 + f_{ap,s_i}(age_i, period_i) + \beta_{bmi} bmi_i + \beta_{unemployment} unemployment_i + \beta_{tax} tax_i + \beta_{alcohol} alcohol_i + \log(E_i) \tag{56}$$

The model defined in equation 56 represents an extension of the original GAM in this thesis and is a departure from the classic APC model. We take advantage of the data enrichment discussed in Subsection 3.2 and integrate additional variables into the model. This is to better represent the differences in mortality rates related to different socioeconomic and health-related factors. The integration of such additional variables in the context of mortality modeling is not a standard approach. It is motivated by Villegas and Haberman (2014), who found that differences, especially those related to socioeconomic conditions, pose a major challenge for the design of public policies to address social inequalities, as well as for the design of pension systems and the management of longevity risk in pension funds and pension portfolios. In contrast, this thesis has mortality data in aggregate form. To the best of our knowledge, such modeling based on the enrichment of publicly available aggregate data has not yet been performed or published. Furthermore, these additional variables contain per-country and per-year values and thus do not provide finer granulation by gender and age. We therefore tend not to view or interpret these covariates as directly influencing variables, but rather as indicators of lifestyle in the country, thus giving the model a more exploratory character.
As described in Subsection 3.2, the selection of the given covariates was influenced by the content-related plausibility aspect. The WHO states that there is a causal relationship between harmful use of alcohol and various mental and behavioural disorders, other noncommunicable conditions as well as injuries (World Health Organization, 2018). According to the European Association for the Study of Obesity statistics, overweight and obesity are the fifth leading risk for global deaths (Czernichow et al., 2021).
It is well-established and evidentially proven by precious studies (e.g. Martikainen and Valkonen, 1996) that higher mortality rates are expected among unemployed people than among those in employment. Tax revenue is included because it has been postulated that by raising more revenue, the state is more likely to be able to invest in health care. This in turn can influence the health status of the population and thus longevity and mortality.

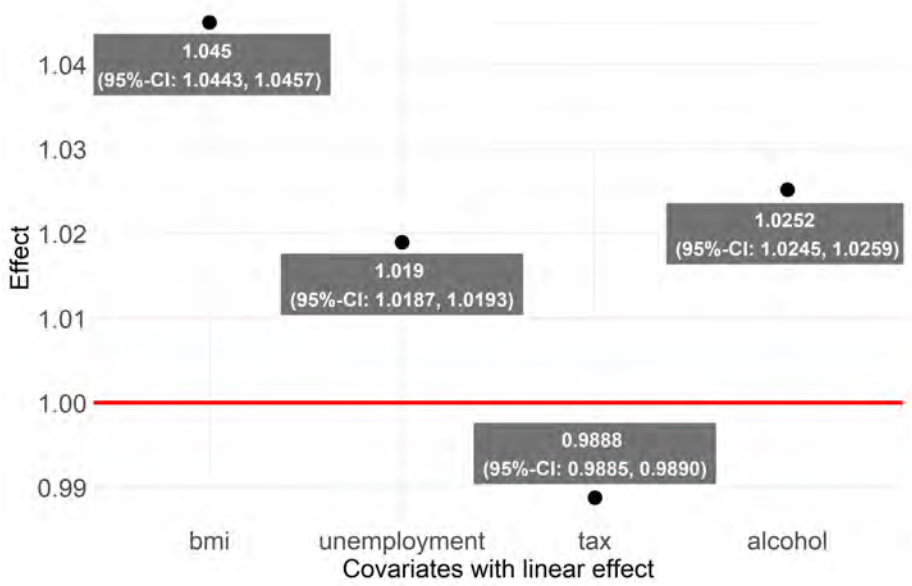


Figure 34: Estimated linear effects of the variables obesity prevalence, unemployment rate, tax revenue and alcohol consumption in the enriched model for the mortality rates. Uncertainty is displayed by 95%-confidence intervals

Figure 34 summarizes the estimated effect as associations between the mortality rates and all included covariates. While the prevalence of overweight measured by BMI, the unemployment rate and the alcohol consumption show a clear positive effect, the mortality rates decrease with increasing tax revenue. More precisely, if the percentage of defined population with a BMI of $25 kg/m^2$ or higher increases by one percentage point, the mean mortality rate increases by $4,5\%$. If the unemployment rate increases by one percentage point, the mean mortality rate increases by $1,9\%$. Further, if alcohol consumption increases by one liter per person, the mean mortality rate increases by $2,52\%$. With one percentage point increase of tax revenue, the mean mortality rate decreases by $1,12\%$. The above interpretation holds under the assumption that everything else, especially age and period (or year) remains the same. Therefore, these effects can be interpreted in the same way for death counts, since it must also be assumed that the exposure, that is, the population size must remain the same. The directions of the effects are plausible. Even if their absolute magnitude appears to be relatively small, this is in line with comparable studies (e.g. Villegas and Haberman, 2014). Moreover, this should be related to the inherently relatively modest values of mortality rates.

Predictive power:

Apart from the interpretation of the effects, it is of high relevance to assess the predictive power of the described models. This is achieved using a test data set that was excluded from the actual training process of the model. We use the forecasts of the 2016-2019 data, where the errors were expressed in RMSE. The table 10 provides a comparison between the different models for each country (row) and gender (column). The individual columns show four models to be compared. First, we mention the classical APC model, which refers to the model introduced in

Subsection 4.1. The second column called GAM represents the APC model built with GAM, which is based on tensor products of age and period estimated for each subpopulations. The third column represents the model enriched with additional socioeconomic and health-risk covariates and its results. Finally, the fourth column is about a continentwise approach. Here, two models were set up that follow the GAM-variant, but were estimated separately for the US. Thus, there is one model for the EU countries and one for the US, which goes hand in hand with the geographical interplay.

Country	Female					Male				
	Class. APC	GAM	Covariates	Continentwise		Class. APC	GAM	Covariates	Continentwise	
FIN	0.0029	0.0012	0.0015	0.0014		0.0029	0.0015	0.0016	0.0015	
DE	0.0046	0.0021	0.0018	0.0019		0.0045	0.002	0.002	0.0023	
ITA	0.0025	0.0016	0.0012	0.0014		0.0021	0.0013	0.0016	0.0015	
NLD	0.0020	0.0013	0.0010	0.0011		0.0038	0.0011	0.0011	0.0012	
US	0.0018	0.0010	0.0014	0.0019		0.0020	0.0016	0.0013	0.0018	

Table 10: Out-of-sample RMSE for all countries, genders and models APC as well as the GAM-based improvements with Age-Period tensor product splines, the integrated socio-economic and health-risk variables and the continentwise modeling. The fitted period is 1990-2015 for Finland, Germany, Italy, the Netherlands and the US. The out-of-sample years are 2016-2019.

When looking at the out-of-sample RMSE, for the classical APC method a major reduction in forecast errors has been achieved by jointly modeling of all subpopulations with GAM. The integration of socio-economic and health variables seems to have a small positive effect on forecast quality. This finding relates to the present case with aggregated data. However, it is worth noting that this modeling and procedure can be applied in exactly the same way to a situation with individual data. It is to be expected that then a possibly higher advantage can be achieved by enriching the model, since then the observations contain specific values for the additional characteristics. This would allow for an analysis at a more granular level. Modeling separated by continents (Europe vs. USA) does not seem to be beneficial. The table of in-sample errors shows similar observations and can be found in the appendix.
In summary, the original assumption that a multipopulational, coherent forecasting method is beneficial and helps the model to learn and adapt the relationships more effectively is validated. So, it lets confirm that considering each country and gender as a membership of a group is preferable to individual modeling and forecasting. This could happen because similar countries in their socio-demographic, economic structure learn from each other and the country-specific characteristics are better distributed among the women and men within the country. Finally, all of the examined countries are highly developed societies that may share many characteristics beyond those of geography.

5.5 Impact of COVID-19 on mortality rates and life expectancy

The focus of this subsection is on dealing with the year 2020 as a breakpoint at the end of the time series, most likely due to COVID-19. For this, the model defined in equation 55 is used as a basis. In Subsection 5.2 we discussed methods for the breakpoint analysis for structural changes being present over time. However, this does not help for modeling COVID, as the structural change is at the end of the time series, thus without future data.
We start with presenting the age, period and cohort effects as before. Figure 35 shows the effects of the model based on the temporal effects age, period, cohort only, which was described as basic or pure in the previous section, since all subpopulations were pooled together. The model here was however trained from 1990 to 2020 inclusive. While the age and cohort effects proceed as expected by the previous model up to the year 2015, the period effect shows a steep increase up to the year 2020, indicating an increased effect of the year on mortality rates. More precisely, the improvement of mortality rates in the last years or even decades was probably just cancelled out by the COVID-19 effect, so that the level of about 2003 is reached again.

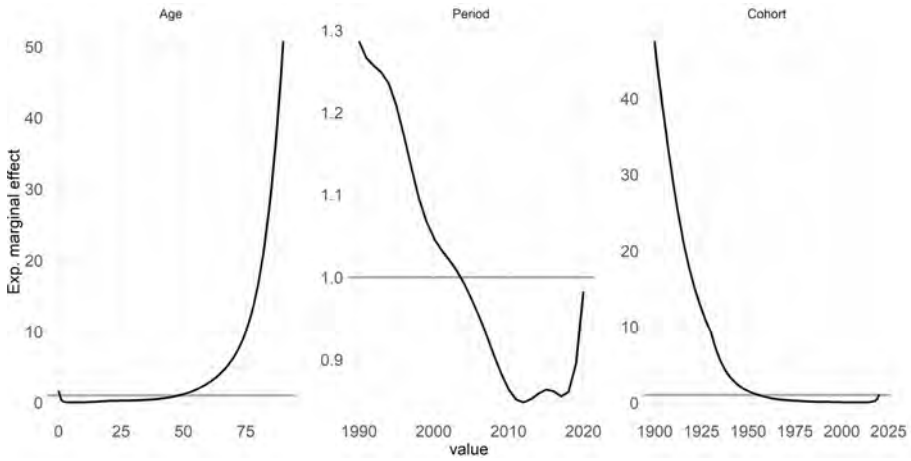


Figure 35: Estimated marginal effects of age, period and cohort for mortality rates. The horizontal lines mark the level of having no effect at all. The GAM was fitted for years 1990-2020, for ages 0-90.

A look at the analysis of the effects broken down by country and gender allows us to understand the background to this occurrence of the period effect (figure 36). When looking at the model outcome including country and gender effects, it seems that there are no changes compared to the model up to 2015 regarding the age and cohort effects. The increase in recent years in the period effect is most clearly visible for US women and men. It is somewhat weaker, but still pronounced, for women and men in Italy as well. These findings may be attributable to the COVID situation in each country. In 2020, incidence and death rates were particularly high in the US and Italy (Johns Hopkins, 2021). In the previous subsection it was already discussed that there is an increased mortality rate in the US, compared to the other countries. Also, other studies show that long before COVID-19, US was at a disadvantage compared to other highly developed countries in terms of increasing mortality. The literature shows that young and middle-aged people were particularly affected. The increasing

mortality in these age groups can be traced back to the deaths that resulted from drug consumption, suicide and cardio-metabolic diseases, other chronic illnesses and injuries (Emanuel et al., 2021). While life expectancy appeared to continue to increase in other countries between 2014 and 2017, mean life expectancy in the US declined to the point that it became the subject of public discussions. Since the pandemic, however, this has been overshadowed by the even larger increase in mortality rates for 2020 (Woolf and Schoomaker, 2019, Bernstein, 2018).

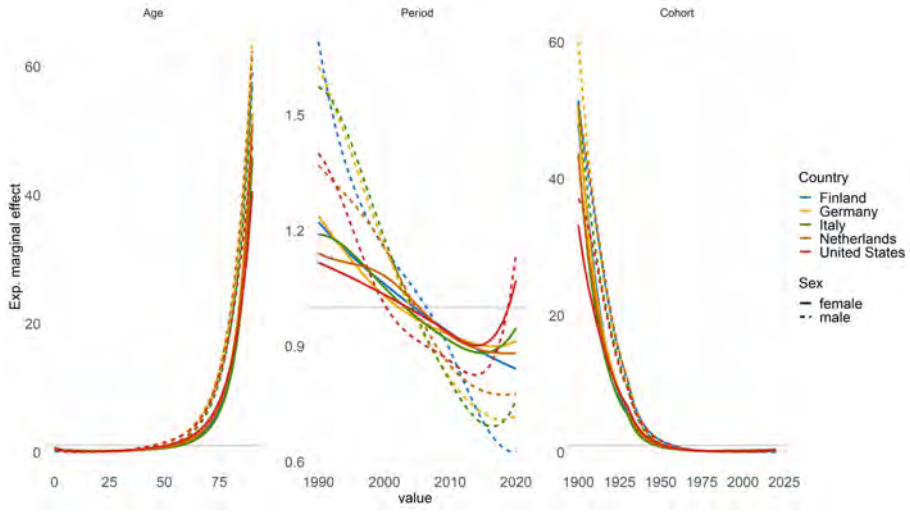


Figure 36: Estimated marginal effects of age, period and cohort for mortality rates, considering the countries and genders in the model. The horizontal lines mark the level of having no effect at all. The GAM was fitted for years 1990-2020, for ages 0-90.

We will now look into the scenarios presented in Subsection 4.4 and the corresponding model outcomes that attempt to outline the future development of COVID-19. In the following, the scenarios are briefly repeated and the assumptions are discussed. Therefore, all results made in this and the subsequent section should be seen as potential examples of a COVID-19 impact analysis. Thus, it is firmly stressed at this point that interpretations of the results should be made in light of the current uncertainties of the further evolution of the pandemic and with all caution. The implementation refers to the model explanations in Subsection 4.4.

For scenario I the the year 2020 is completely omitted from the model training process, so that only data up to 2019 are included.

Scenario II assumes that the country-specific COVID-effect will continue in the same way in the next several years. Technically, this means that the model was fitted with data up to 2020 and the COVID-indicator was set to one for the next years.

Scenario III, on the other hand, assumes that the COVID-effect will be present in the next coming years, but will flatten exponentially over time. This was implemented by assigning exponentially decreasing values between one and zero for the COVID-indicator for the following years.

For the adjustment for excess mortality in the fourth scenario, the actual numbers of deaths in 2020 were first contrasted with the number of deaths expected based on the model from scenario I, which are not influenced by the pandemic.

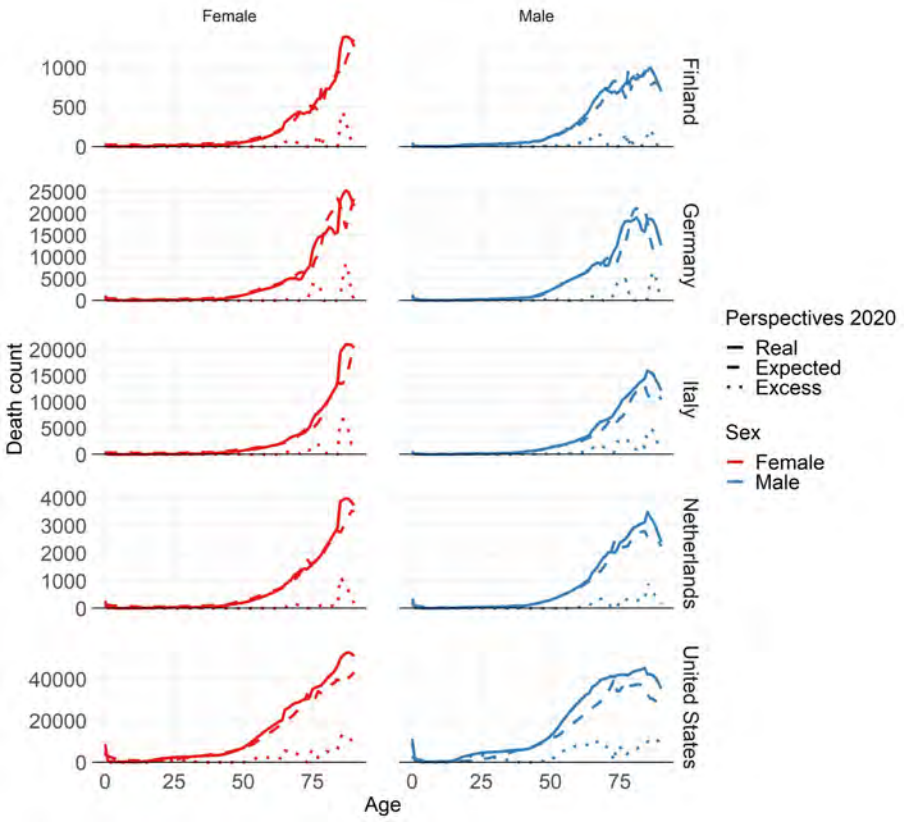


Figure 37: Excess mortality for year 2020. Here the expected and actual mortalities for all subpopulations are contrasted.

Figure 37 shows the calculated excess deaths. While the solid line represents the actual death counts in 2020, the dashed lines represent the expected death counts for that year. The dotted lines below are the positive deviations used for modeling, hence the excess mortality that is adjusted from the 2021 population for modeling scenario IV. Visually, the comparison between countries can only be made in relation to population counts, as the absolute death figures are presented here. To train the model, the death rates for the years 2020 and 2021 are used based on the prediction of the first scenario, i.e. without the influence of COVID. The population for 2021 is calculated by explicitly adjusting for and thus removing the excess mortality in 2020, whereby the excess mortality is based on the calculation described above. The underlying assumption here is that older people who were particularly vulnerable to the virus have already died off and that correspondingly fewer in these cohorts will die in the coming years.

As a reminder, the GAMs are defined by a two-dimensional cubic smoothing splines, called tensor product. In general, the two-dimensional smooth functions are implemented by the so-called penalized or P-splines, for which further details can be found either in Subsection 4.3 or Wood (2011, 2017). Here, predictions of future mortality rates are based on an extrapolation of the spline fit, with the penalized smoothing second-order spline bases assuming a globally quadratic structure and a persistent curvature outside the observed data. Consequently, the predictions of future mortality rates depend also on the degrees of freedom chosen for the covariates or estimated automatically as in this case. The out-of-data forecasts should be made cautiously, as the authors recommend that it is prudent not to extrapolate far into

the future (Wood, 2017). We will therefore limit the analysis on the trend forecast to the year 2025 at the longest. It is also suggested that when using GAMs around past trends for extrapolation purposes, the degrees of freedom should be chosen as much as feasible to catch the minimum necessary trends in order to ignore random fluctuations in the data. The reduction of the number of nodes in both dimensions to five can help to get this challenge under control.

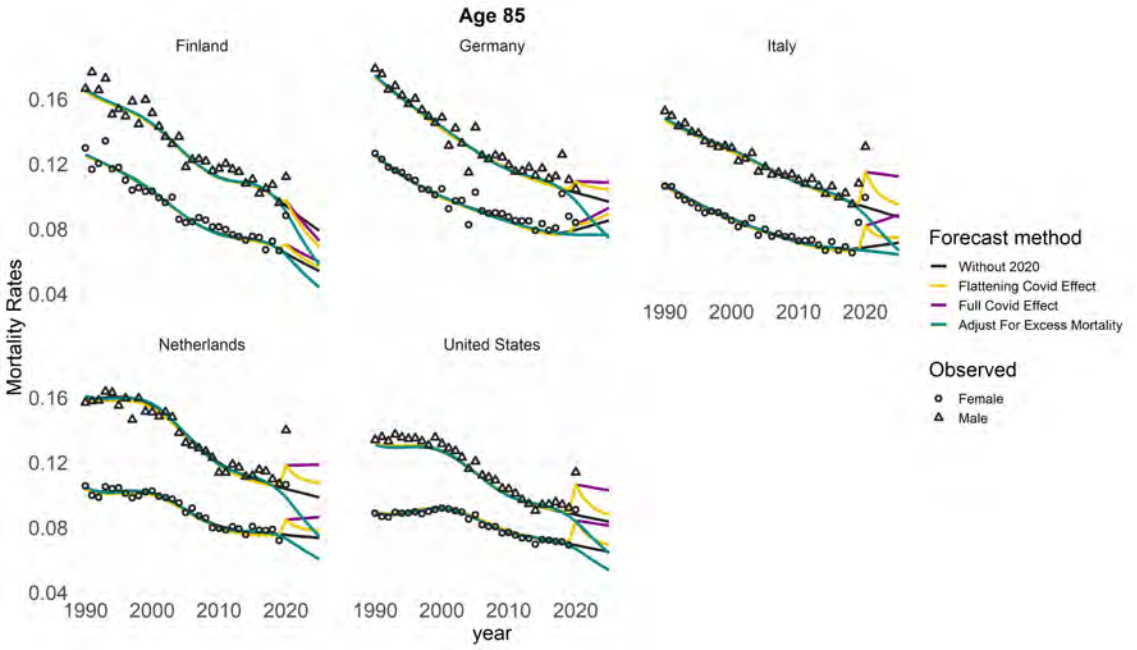


Figure 38: Different trend forecasts according to the scenarios for 85-year-olds. The adjusted years are 1990-2020 and 1990-2021 for the scenario demanding an adjustment for excess mortality. The forecast extends to the year 2025.

Figure 38 illustrates the outcomes of our four scenarios to the considered countries for males and females exemplarily for 85 years old people. The black circles and triangles show the observed mortality rates for females and males, respectively. Comparing the observations with the alignment of the line, we can speak of an overall good fit by visual inspection, as the estimated mortality rates are close to the observed mortality rates, while the curve is still smooth. We deliberately do not want to achieve an overfit, in order not to project the random noise in the data along with it into the future. For this age catogory, for which the COVID-effect is supposed to be quite high from a medical and epidomiological point of view, an outlier can be clearly seen especially for Italy, US and Netherlands, as the corresponding point tears very far upwards. For Finland and Germany, on the other hand, there is hardly any particularity to be seen for year 2020. The interpretation of the content of the individual trend forecasts therefore seems to work conclusively, especially for the subpopulations strongly affected by COVID. After all, our scenarios involving the COVID-effect are based on the supposition that this effect exists in the populations at hand. Thus, the black curve can be seen as a fairly optimistic continuation of the trend, while the second scenario expects the highest mortality rates via the constant COVID-effect, marked by the purple line. On the other hand, the yellow trend, which assumes that the COVID-effect flattens with time, finds its origin in the same point where the purple line is, but becomes increasingly milder with time

and decreases exponentially, as expected. In addition, the turquoise prediction is even more optimistic, so it is even below the black line. This means that we expect the lowest mortality rates in this scenario, which can be logically combined with the underlying idea that we do not expect any more COVID-deaths in the future and instead clean up the population on top of those who have already died in 2020. For example, in the case of Italy, it is striking that the predicted curve for the fourth scenario slopes more steeply for men than for women. This is due to the greater excess mortality of men in this country. Of course, this also means that the populations that were most affected by the outlier in 2020 are predicted to have an even lower trajectory. Men in Finland represent a small positive deviation from this description. This can be explained by the fairly good COVID-handling in Finland, so that COVID-related effects are generally not very pronounced here. Moreover, the result is probably influenced by the past trend of mortality rates in this country, which shows a downward slope.

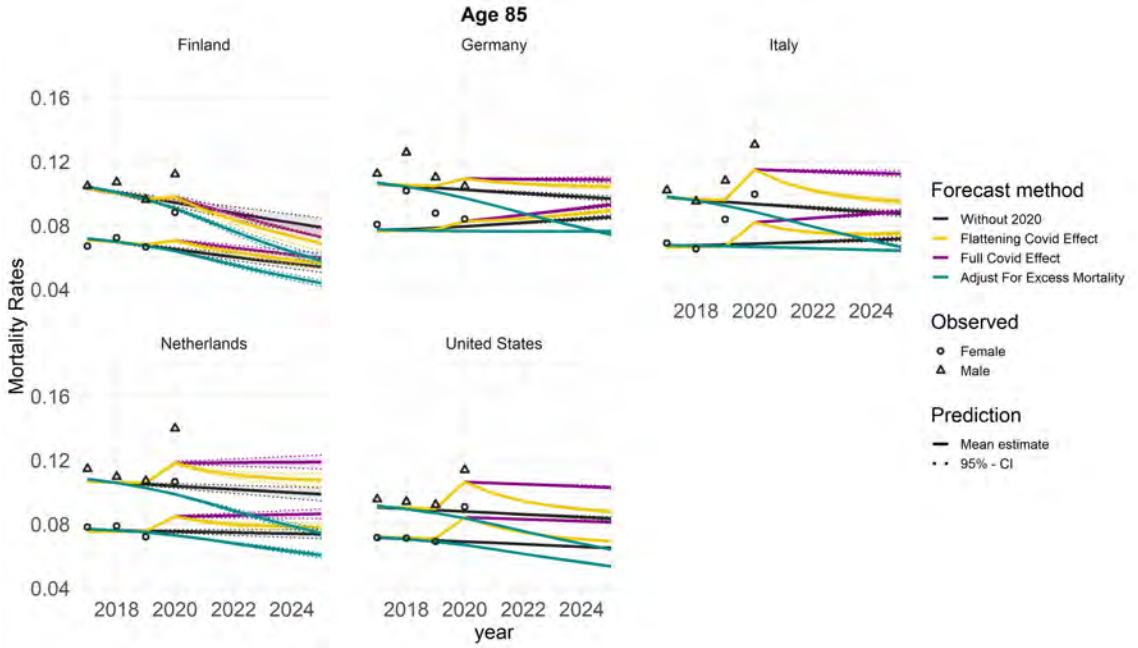


Figure 39: Different trend forecasts according to the scenarios for 85-year-olds. The adjusted years are 1990-2020 and 1990-2021 for the scenario demanding an adjustment for excess mortality. The forecast extends to the year 2025. Dashed segments represent the model-based 95%-confidence intervals.

A detailed view is provided by the figure 39 by restricting the annual axis to the years 2017 to 2025. In addition, the corresponding 95% confidence intervals are shown here. These were corrected by the application of Delta method. This method provides a basis for calculating the standard error for transformations. In general, the variance of a transformation $g(.)$ is not equal to the transformation of the variance. Especially for a regression with log-transformed response, the confidence interval for the estimate of μ on the original scale cannot be calculated with $\exp(\sigma)$. In this case, θ represents the mean. Instead, the delta rule must be applied. This states that if $\mu \sim \mathcal{N}(\theta, \sigma^2)$, then the distribution of $g(\mu)$ based on a Taylor-approximation can be locally approximated by $g(\mu) \sim \mathcal{N}(g(\theta), \sigma^2 g'(\theta)^2)$ (Casella and Berger, 2002).

Almost everywhere, the confidence bands are very narrow. One exception is the already conspicuous subpopulation of Finnish men. These have comparatively wider confidence bands, indicating a higher uncertainty of the estimate for this subpopulation, which in turn is in line with the above-mentioned noticeable phenomenon. Visually, one could argue that these overly narrow, model-based confidence bands are overly optimistic and should therefore be taken advisedly.
In the context of forecasting into the future, it is naturally not possible to make an exact quantitative evaluation of the forecast quality, since the actual mortality rates are not yet known. However, in order to be able to assess the plausibility of the forecast, the various trends of the forecast are discussed with regard to the underlying assumptions of the individual scenarios. The different assumptions underlying the four illustrative alternative scenarios imply a range from mild to severe expected impacts and will be discussed in detail in Section 6.

Life expectancy:

In some cases, it may be advantageous to express age-specific mortality rates in terms of life expectancy. Life expectancy is one of the most important demographic indicators for comparing different population groups against each other in terms of longevity and represents one of the main outcomes of lifetable analysis. Life expectancy compresses the mean number of years a person in a population is expected to live, after the a^{th} year of life, and can also be understood as the area under the survival curve regardless of its shape. In general, this ratio is often used to characterize life expectancy at birth, i.e. at age 0. However, we will consider here the higher age categories 50-90, as they are potentially more influenced by the COVID-19 pandemic, especially when we focus on the substantive and policy analyses, from the perspective of the need for Medicare and Social Security. In addition, longevity is typically seen at the edge of the mortality curve, as it may be afflicted by non-standard fluctuations at the beginning. In this section, we will convert the previously modeled trend forecasts of age-specific mortality rates into life expectancy for different subpopulations and present them in tabular form for reference.
Let the life expectancy of a person of age a in year t be defined as $e_{a,t}$, the number of years we expect this person to live if we know that this person has survived to age a. To put it in other words: We assume that this person will live an average of $a + e_{a,t}$ years. For the calculation one needs the probability $p_{a,t,k}$ that this person will survive $a + k$ years. This can be calculated as follows:

$$p_{a,t,k} = (1 - \mu_{a,t})(1 - \mu_{a+1,t+1}) \cdots (1 - \mu_{a+k,t+k}) = \prod_{i=0}^{k} (1 - \mu_{a+i,t+i}) \tag{57}$$

Using the representation

$$\mathbb{P}(T_{a,t} > k) = p_{a,t,k} \tag{58}$$

one can express T as the remaining lifetime of that person aged a in year t. With this, $\mathbb{P}(T_{a,t} = k)$ denotes the probability that the individual survives to age $a + k - 1$

and dies in year $t + k$ at age $a + k$. With this notation, life expectancy is derived from mortality rates as follows:

$$e_{a,t} = \mathbb{E}(T_{a,t}) = \sum_{k=1}^{\infty} k \; \mathbb{P}(T_{a,t} = k) = \sum_{k=1}^{\infty} k \; p_{a,t,k-1} \; \mu_{a+k,t+k} \tag{59}$$

This methodology was implemented using the R-package `Mortalitylaws` (Pascariu, 2018). Below we present the tables reflecting the predicted life expectancies of US men 11 and women 12.

The first column shows the corresponding age category and includes ages 50 to 90. The second column shows life expectancies from 2019 derived from actual observed mortality rates. The predictions based on the trend forecasts of mortality rates for the years 2022 and 2023 presented in the previous section were derived using forecasts from both scenarios. In addition, the following columns show the raw differences in age-specific life expectancy in years that would result if the COVID-effect were assumed to remain present. The last two columns show the percentage drop in life expectancy when moving from a COVID-free forecast to a forecast taking 2020 into account. While the raw differences reach up to a maximum of about 1.5 years and tend to be largest for younger age categories, the largest percentage drops tend to be seen between ages $80 - 86$ and range around about 6%. The tables for other subpopulations can be found in the appendix.

Age	2019	Scenario I		Scenario II		Difference		Drop in %	
		2022	2023	2022	2023	2022	2023	2022	2023
50	29.30	29.69	29.72	28.16	28.18	1.53	1.54	5.15	5.18
51	28.44	28.87	28.90	27.36	27.38	1.51	1.52	5.23	5.26
52	27.59	28.05	28.08	26.56	26.58	1.49	1.50	5.31	5.34
53	26.75	27.23	27.27	25.76	25.79	1.47	1.48	5.40	5.43
54	25.91	26.41	26.45	24.97	25.01	1.44	1.44	5.45	5.44
55	25.09	25.60	25.64	24.19	24.23	1.41	1.41	5.51	5.50
56	24.28	24.78	24.83	23.41	23.46	1.37	1.37	5.53	5.52
57	23.48	23.97	24.02	22.64	22.69	1.33	1.33	5.55	5.54
58	22.69	23.17	23.22	21.87	21.93	1.30	1.29	5.61	5.56
59	21.91	22.36	22.41	21.11	21.17	1.25	1.24	5.59	5.53
60	21.14	21.57	21.61	20.36	20.42	1.21	1.19	5.61	5.51
61	20.37	20.77	20.82	19.61	19.68	1.16	1.14	5.58	5.48
62	19.62	19.98	20.03	18.87	18.94	1.11	1.09	5.56	5.44
63	18.88	19.20	19.25	18.14	18.21	1.06	1.04	5.52	5.40
64	18.15	18.43	18.47	17.41	17.49	1.02	0.98	5.53	5.31
65	17.42	17.66	17.70	16.70	16.77	0.96	0.93	5.44	5.25
66	16.70	16.90	16.94	15.98	16.06	0.92	0.88	5.44	5.19
67	15.98	16.15	16.19	15.28	15.36	0.87	0.83	5.39	5.13
68	15.27	15.41	15.44	14.59	14.67	0.82	0.77	5.32	4.99
69	14.56	14.68	14.71	13.90	13.98	0.78	0.73	5.31	4.96
70	13.86	13.96	13.99	13.22	13.30	0.74	0.69	5.30	4.93
71	13.16	13.25	13.28	12.56	12.63	0.69	0.65	5.21	4.89
72	12.47	12.55	12.58	11.90	11.97	0.65	0.61	5.18	4.85
73	11.79	11.87	11.89	11.25	11.31	0.62	0.58	5.22	4.88
74	11.11	11.19	11.21	10.60	10.67	0.59	0.54	5.27	4.82
75	10.45	10.53	10.55	9.97	10.03	0.56	0.52	5.32	4.93
76	9.80	9.88	9.90	9.35	9.40	0.53	0.50	5.36	5.05
77	9.16	9.24	9.26	8.73	8.78	0.51	0.48	5.52	5.18
78	8.52	8.61	8.63	8.13	8.17	0.48	0.46	5.57	5.33
79	7.89	7.99	8.01	7.53	7.57	0.46	0.44	5.76	5.49
80	7.27	7.37	7.39	6.94	6.97	0.43	0.42	5.83	5.68
81	6.65	6.76	6.78	6.35	6.37	0.41	0.41	6.07	6.05
82	6.03	6.15	6.17	5.77	5.79	0.38	0.38	6.18	6.16
83	5.42	5.54	5.55	5.19	5.20	0.35	0.35	6.32	6.31
84	4.80	4.91	4.93	4.60	4.61	0.31	0.32	6.31	6.49
85	4.17	4.28	4.29	4.01	4.01	0.27	0.28	6.31	6.53
86	3.52	3.62	3.63	3.40	3.39	0.22	0.24	6.08	6.61
87	2.85	2.92	2.93	2.76	2.76	0.16	0.17	5.48	5.80
88	2.14	2.18	2.19	2.08	2.08	0.10	0.11	4.59	5.02
89	1.36	1.38	1.38	1.33	1.33	0.05	0.05	3.62	3.62
90	0.49	0.49	0.49	0.49	0.49	0.00	0.00	0.00	0.00

Table 11: Life expectancy forecasts for US male population

Age	2019	Scenario I		Scenario II		Difference		Drop in %	
		2022	2023	2022	2023	2022	2023	2022	2023
50	32.29	32.38	32.46	31.47	31.66	0.91	0.80	2.81	2.46
51	31.38	31.49	31.57	30.60	30.78	0.89	0.79	2.83	2.50
52	30.48	30.60	30.68	29.72	29.90	0.88	0.78	2.88	2.54
53	29.59	29.72	29.79	28.85	29.03	0.87	0.76	2.93	2.55
54	28.70	28.84	28.91	27.99	28.16	0.85	0.75	2.95	2.59
55	27.82	27.96	28.04	27.12	27.30	0.84	0.74	3.00	2.64
56	26.95	27.09	27.16	26.27	26.44	0.82	0.72	3.03	2.65
57	26.08	26.22	26.29	25.41	25.58	0.81	0.71	3.09	2.70
58	25.22	25.35	25.42	24.56	24.72	0.79	0.70	3.12	2.75
59	24.37	24.49	24.56	23.71	23.87	0.78	0.69	3.18	2.81
60	23.52	23.63	23.70	22.87	23.03	0.76	0.67	3.22	2.83
61	22.68	22.78	22.85	22.03	22.19	0.75	0.66	3.29	2.89
62	21.85	21.93	22.00	21.20	21.35	0.73	0.65	3.33	2.95
63	21.02	21.09	21.16	20.38	20.52	0.71	0.64	3.37	3.02
64	20.19	20.26	20.32	19.56	19.69	0.70	0.63	3.46	3.10
65	19.37	19.43	19.49	18.74	18.87	0.69	0.62	3.55	3.18
66	18.56	18.61	18.67	17.94	18.06	0.67	0.61	3.60	3.27
67	17.74	17.79	17.85	17.14	17.26	0.65	0.59	3.65	3.31
68	16.94	16.99	17.04	16.35	16.46	0.64	0.58	3.77	3.40
69	16.14	16.19	16.24	15.56	15.67	0.63	0.57	3.89	3.51
70	15.34	15.40	15.45	14.79	14.89	0.61	0.56	3.96	3.62
71	14.56	14.62	14.66	14.02	14.11	0.60	0.55	4.10	3.75
72	13.79	13.84	13.89	13.26	13.35	0.58	0.54	4.19	3.89
73	13.03	13.08	13.12	12.52	12.59	0.56	0.53	4.28	4.04
74	12.27	12.32	12.36	11.78	11.85	0.54	0.51	4.38	4.13
75	11.52	11.58	11.61	11.05	11.11	0.53	0.50	4.58	4.31
76	10.79	10.84	10.88	10.33	10.39	0.51	0.49	4.70	4.50
77	10.06	10.11	10.14	9.63	9.67	0.48	0.47	4.75	4.64
78	9.34	9.39	9.42	8.93	8.97	0.46	0.45	4.90	4.78
79	8.63	8.68	8.71	8.24	8.28	0.44	0.43	5.07	4.94
80	7.92	7.97	8.00	7.56	7.59	0.41	0.41	5.14	5.13
81	7.22	7.27	7.30	6.89	6.91	0.38	0.39	5.23	5.34
82	6.53	6.58	6.60	6.23	6.24	0.35	0.36	5.32	5.45
83	5.83	5.88	5.90	5.57	5.58	0.31	0.32	5.27	5.42
84	5.14	5.18	5.19	4.91	4.91	0.27	0.28	5.21	5.39
85	4.44	4.47	4.48	4.25	4.25	0.22	0.23	4.92	5.13
86	3.72	3.75	3.76	3.57	3.57	0.18	0.19	4.80	5.05
87	2.98	3.00	3.01	2.87	2.87	0.13	0.14	4.33	4.65
88	2.21	2.22	2.22	2.14	2.14	0.08	0.08	3.60	3.60
89	1.39	1.39	1.39	1.36	1.36	0.03	0.03	2.16	2.16
90	0.49	0.49	0.49	0.49	0.49	0.00	0.00	0.00	0.00

Table 12: Life expectancy forecasts for US female population

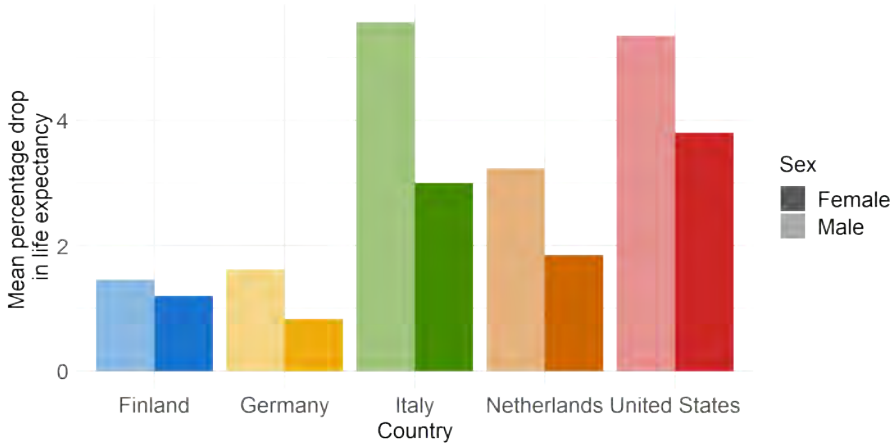


Figure 40: Mean percentage drop in life expectancy forecast for year 2022 by country and gender

If one compares the differences in the forecasts for the life expectancy of 50- to 90-year-olds between Scenarios I and II, a decline in life expectancy for all populations can be observed. In summary, the figure 40 draws a comparison between the two scenarios in terms of the decline in life expectancy due to the COVID situation. The horizontal axis represents the averaged value of the percentage drop for the forecast of the year 2022. The forecasts show the highest decline in life expectancy for the US and Italian populations, especially for males, followed by females. In the case of Italy, the COVID-19 effect may be due to the health care system, which was possibly overloaded and may, with a rapid response and decent preparation, be a one-time effect. In contrast, the situation in the US seems to be related to the lack of public health insurance and is not a unique event. The Netherlands also seems to be particularly affected by the addition of the COVID-19 effect in 2020 and to carry over the effects into the future to a relatively high degree. In contrast, Finland and Germany appear to lose comparatively few life years to the COVID-situation in the future. These findings are consistent with the results from previous sections as well as the research results of previous studies and numerical data (Johns Hopkins, 2021).

6 Discussion

In this thesis, we explore alternatives and improvements to existing mortality modeling and prediction capabilities. Particularly, we improve the classical Lee Carter and APC models with machine learning and GAM, respectively, and analyse the impact of the pandemic on mortality rates with different scenarios. In addition, the impact of COVID-19 on the trends forecasts of mortality rates as well as life expectancy is analyzed and estimated for the US and compared to other populations.

Data:

All of these analyses are based primarily on open source data provided by Human Mortality Database (HMD, 2021). This data source provides high quality data by following numerous criteria until individual countries are allowed to put their data on the website. In particular, only mortality data from a sound and statistically highly

qualified source is allowed to be included, for example, those that ensure almost one hundred percent population coverage based on census and vital registration systems. For the enrichment of recent years the STMF dataset was used additionally, which is necessary to examine the COVID-19 impact (STMF, 2021). This provides up-to-date mortality data on a weekly basis. However, pooling the data into relatively broad age groups and mapping them to the metric scale only in retrospect could cause biased results, since the assumption that the distribution of death rates within the given age groups remains the same could be violated.
Another item is the data enrichment with socio-economic and health variables. It is about applying the models discussed earlier also to study the relationship between socio-economic or health circumstances with mortality. Although this results in a useful exploratory area of investigation, the application of such specifications in the context of aggregate data is subject to more detailed examination. While it would be desirable to have the data at the same level of aggregation, the values of the individual indicators are not age- and gender-specific, whereas the mortality rates are. Nevertheless, one can suppose that the covariates considered in this thesis should be seen more as an indicator of country-specific characteristics and are assumed to have similiar impact on the health and socioeconomic situation of the entire population. In any case, the methodology can be applied in the same form to a set of individual data and is certainly an attractive field for further research.

Modeling:

The modeling process needs to be carefully examined when fitting time series to period effects for a long period of time. We find out that the presence of structural breaks in the period effects is significant but not conclusive. The fact that structural breaks occur more frequently in the male data than in the female data suggests that the improvement in mortality rates has changed more for males than for females, at least during the period considered in this thesis. When considering the period effect, it is important to note that a visual inspection of a graph of $\kappa_{t,s}$ shows us that each change in improvement has accelerated the improvement in mortality rates above those of previous years. Put another way, if this perception is not included in the model because of the time span considered in the data used for this work, there is a risk that mortality rates will be overestimated. It is also worth noting that in the course of the analysis we realised that the CBD and M7 models are more suitable for older age categories.
It can be established that generalized additive model is superior and allows for multipopulational modelling. Regarding the modeling with GAMs, a notable limitation of all projections made with GAMs is their dependence on the shape of the models, viewed as an exercise in curve fitting with extrapolation. The validity of each projection depends on whether the model class follows the expected changes over time. In our particular case, a globally quadratic structure is assumed based on the penalized smoothing of B-splines, so that continuous curvature is supposed outside the observed data.

COVID-Scenarios:

In the course of predicting future mortality and life expectancy trends, various scenarios have been set up to counteract the considerable uncertainty of the current COVID-19 situation. However, these scenarios are bound up with assumptions that can be discussed extensively.

The first scenario is about the idea that COVID-19 would completely disappear in the future. Thus, we not only assume that the resulting fluctuations seen in 2020 are entirely due to the pandemic, but also that it is a residual event that will not affect mortality at all in the coming years. On the other hand, it must be noted that at the time of writing this thesis, there are no evidence-based studies on the long-term effects that would provide one hundred percent certainty that the health of the recovered will not be at risk at all in the long run. Rather, the phenomenon of "Long Covid" is increasingly being reported by those who have already recovered. Even people report exhaustion, fatigue and reduced performance, in whom the acute illness was not severe at all. These late effects of the disease are collectively referred to as "Long Covid". Mental illnesses also occur (Sudre et al., 2021). The nowadays widespread and massively used vaccinations against COVID-19 give hope that the situation can be kept under control in the future, but this gives no guarantee how fast and to what extent this will lead to an abrupt disappearance of the pandemic. In addition, the database on the existence and consequences of long-term side effects of the individual vaccines is not yet reliable enough to be able to make forecasts in this regard, according to the current state of knowledge. In some countries, lockdowns lasting several months were used in 2020 as one of the few measures perceived as effective by the political authorities to combat the pandemic. This can be a psychological burden in the long run, as LMU Psychiatry was able to study in its patients (Kunzler et al., 2021). Long-term effects may also be expected in this regard. Furthermore, it is implicitly assumed that the excess mortality will average out over the next few years, so that no excessive, rapid reductions in the population are to be expected. Whether this is really the case, and whether the assumption that no long-term effects on health are to be expected is justified, is the subject of discussion and certainly an exciting field of investigation for future generations of research.

According to the second scenario, we expect a fully continuing effect of COVID-19 on the mortality rates in the future. On the one hand, we assume that the pandemic-related situation will continue in the same way as in 2020, which can be considered as an extreme assumption. As already experienced in recent months, scientific and medical progress in this field is not idle. Due to the vaccination campaigns used in all countries covered in this thesis, this assumption is already violated to the extent that, at least as the studies show nowadays, the risk of infection is reduced for a broad mass of the population (Polack et al., 2020). However, current evidence from various countries seems to show that despite a large proportion of those vaccinated, incidence rates continue to rise, sometimes exponentially (Johns Hopkins, 2021). Further, we assume that the impact of COVID-19 on mortality will remain exactly the same over the next five years. This is currently difficult to assess, as we can only speak of short-term decrease in the death rates (Boudourakis and Uppal, 2021). Apart from that, this decrease is questionable because no exact backward mapping can be made to the cause of these changes. It is not completely transparent whether these changes are of a seasonal nature or can actually be attributed to the vaccinations and thus

rule out the return of the higher death rates. Conversely, the various emerging viral variants may call into question the effectiveness of the vaccines. This could be a barrier to improving the pandemic-related situation. In the worst case, lockdowns could be reinstated. These are not only fraught with negative economic consequences, but may rather further exacerbate the health consequences. The extent to which such measures will influence the COVID-19 effect on health in the long term cannot be estimated with certainty. It is often argued that a prolonged strict lockdown has limited success in containing the spread of the virus, so even after some relaxation, containment measures must be implemented. Continued efforts to prevent the spread of the virus would continue to significantly dampen economic activity in all sectors until a vaccine (or other effective medical solution) becomes available. Overall, this scenario envisions significant and lasting impacts on health-related sectors.

The third scenario, which suggests that the COVID-effect will level off over the years, seems to represent an optimistic and in some respects possibly even more realistic development. As with scenario 2, we assume that 2020 is not an outlier that may be completely disregarded, but we give the relevance for the future a reduced intensity. The basic assumption here is that the pandemic will have progressively less impact on health and mortality and will almost completely disappear after five years. This represents a kind of mixture of two chains of reasoning from the previous scenarios. Because on the one hand we assume that the pandemic will influence the residual effect. However, at the same time we recognize that, among other things, medical progress, changes in people's behavior, and possibly herd or vaccine immunity, which will build up more and more every year, will make this effect increasingly irrelevant.

In the course of the fourth scenario, the focus is placed on the fact that the excess mortality from the year 2020 is to be adjusted from the following years. Some of the assumptions are based on Scenario I, because we again assume that the adverse health effect of pandemic will disappear and no long-term consequences will happen. In contrast to the above scenario, however, we additionally assume that excess mortality must be explicitly taken into account because it will not average out in the next few years. To do justice to this assumption, we have to assume that baseline mortality will not change in the next few years, so that the surpluses from 2020 will become noticeable in this way. Thus, it is implicit to assume that there were no behavioral changes in all age categories as a result of the pandemic. However, this assumption seems highly questionable. For example, a large proportion of the working population had to work mostly from home during the peak periods of the pandemic. As a result, the commute to work was eliminated. This in turn could be a reason for fewer traffic fatalities, which of course directly affects the mortality trend expressed in the final numbers. Vacations were also limited and in some cases impossible due to travel bans. This could have resulted in less skiing and accidental deaths. Finally, quarantine measures, excessive attention to hygiene, and contact restrictions may have led to fewer influenza deaths. These are all possible reasons that argue against the basic mortality being absolutely unaffected by the pandemic.

As can be seen by the above reasoning, the establishment of possible COVID-related future scenarios is not a self-evident task. The risk arising from the pandemic depends on numerous factors, including political decisions and social acceptance of them. Yet, to provide mortality forecasts also with regard to the future attitudes of life insurers, the presented scenarios are a helpful and valuable basis.

7 Conclusion

Mortality forecasts are important for predicting the future extent of population aging and for establishing the sustainability of pension and social security systems. It has become even more relevant due to the current events caused by the COVID-19 pandemic, emphasizing that the problem we tackled in this thesis is very up-to-date. Also, the comparison of the results using data from HMD of women and men from five countries is the subject of the investigation: Finland, Germany, Italy, the Netherlands and the US.
Before starting the main analysis, a data enrichment process was carried out to map the mortality data of missing years from grouped to metric age scale. On the other hand, additional covariates were integrated from secondary data sources.
To summarize the results obtained in this thesis, we will revisit the research questions posed in the introduction. We have examined various methods to model the mortality rates of a person with a certain age in a certain year to find out how the state-of-the-art mortality models can be improved in terms of fit and forecast. Starting with the Lee Carter model under Poisson setting, we not only compared this with three other extensions, but also performed a detection of structured breaks. The breakpoint analysis proves to be useful and is identified for each of the following three subpopulations: Italian (1983), Dutch (1993), US (1968) male populations. It is likely that these years can be attributed as decisive to health-care and health-assurance reforms and policies. The fact that a breakpoint is identified rather for men than for women is in line with the findings of other studies and suggests that the improvement in mortality rates over the last decades has been more rapid for men. This could also be related to the narrowing gap between male and female mortality rates over the last years, especially in the older age categories, although in general higher mortality rates for males than females can still be observed in all nations. Since both the in-sample and the out-of-sample prediction quality is improved by this, the decision was made to take the detected structural changes into account, wherever possible, in the further modeling procedure. The model proposed by (Lee and Carter, 1992) seems to be widely accepted for providing satisfactory adjustments and forecasts of mortality rates. This method is therefore widely used, and several extensions and modifications have been proposed to achieve a broader interpretation to capture the main characteristics of the dynamics of mortality rates, and thus to improve the performance of the Lee Carter model. The other extensions that were applied and compared in this thesis are the CBD model (Cairns et al., 2006) and the so-called Quadratic CBD model (M7) (Cairns et al., 2009) as well as the APC model (Clayton and Schifflers, 1987, Hobcraft et al., 1982, Holford, 1983, Currie, 2006). Of the classic models, LC and APC outperform the others. Especially for the lower age categories, CBD and M7 are comparatively less suitable.
After comparing the models and crystallizing both as the best performing, the first task is to exploit their strengths and simultaneously further develop and improve them with additional methods that may be less established in the context of demography. To this end, the focus is on evaluating the contribution of machine learning methods to mortality modeling, which primarily seek to adjust what is probably the most important drawback of the Lee Carter model. While the LC model attempts to represent the mortality surface as accurately as possible and also to anticipate future changes in mortality structures, it assumes that the age-varying index beta is

fixed over time. In developed countries, this assumption is violated in any case, as the slowdown in mortality is caused by a decline at younger ages (infant and child mortality) and an acceleration at older ages (Lee and Miller, 2001, Girosi and King, 2008). Therefore, we use tree-based machine learning techniques to address this problem, to use them as a diagnostic tool to detect the weaknesses of the LC model and to enhance it by using an improvement factor (Deprez et al., 2017, Levantesi and Pizzorusso, 2019). We can state that the contribution of machine learning methods to the improvement of fit is undeniable. Of the three tree-based methods Decision Tree, Random Forest and Gradient Boosting Machine models, GBM works best. However, the analysis of goodness-of-forecast shows minimal to no differences, noteworthy for Random Forest. Presumably, this fact is due to the fact that the tree-based models provide better resolution locally. Lee Carter, instead, fits an overall linear trend, which is suitable for global projection into the future. Investigating further methods based on this to boost forecast performance is certainly an interesting aspect of further studies.
Further, the APC model is revisited and modeled by a generalized additive model with a two-dimensional tensor product spline. This does not only provide an intuitively simple method to model the past and to project it into the future, but also allows for extensive interpretability of individual components of age, period and cohort, among others, by omitting constraints. This represents an alternative multipopulational strategy in contrast to the subpopulation-specific models which were presented so far. This model achieves significantly better prediction quality than the classical APC model, both in terms of in-sample and out-of-sample errors. Supposedly, this happens because we allow, through coherent modeling, not only genders within a country, but also geographically, even politically contiguous countries to learn from each other. Another focus is on the feasibility of including additional covariates in the model, as suggested by GAM, and thus, analyzing potential discrepancies in socioeconomic and health-related differences in mortality in a more granular way. This primarily helps the model to properly capture country effects. Nevertheless, the developed model can be applied to individual data, as well.
Having achieved significant advantages for the APC model through GAM in the course of the extensive model comparison, it is natural that we use this to analyze the second research question, namely how mortality will develop in the future in different countries, taking into account the COVID-19 impact. In this thesis, we have discussed current topics that will continue to be important in the future, as predictions of longevity are likely to keep changing over time. The prediction and transferability of the models into the future are difficult to assess, that is why we are looking at potential scenarios. On the one hand, there is naturally no validation reference for this and, on the other hand, the current situation of COVID-19 is subject to uncertainty from both a medical and epidemiological perspective. Therefore, we have set up and discussed different mild to severe scenarios, to make it possible to prepare for future evolutions. The established four scenarios are used to illustrate possible developments, the quality of which can only be assessed years later. For this purpose, we provide trend forecasts for mortality rates and, based on them, for age-specific life expectancies, for women and men in all five studied countries. It is evident that in 2022, the largest average drops in life expectancy due to COVID-19 are expected for the US and Italian male populations, up to more than 6%, especially for the age groups 80-86. The least affected populations are likely to be those in

Germany and Finland. In this thesis, we compared different models to each other. It would be interesting to extend the investigation to Deep Learning techniques and to analyze the differences. Based on the representation learning paradigm, Richman et al. (2020) propose to extend the Lee Carter model to multiple populations using neural networks. These should automatically select an optimal model structure. It is claimed that the out-of-sample prediction performance of the model is very competitive. It would be of enormous interest to apply this model to the data used so far in this thesis and, in particular, to compare it with the APC method patterned via GAM.
The challenge in practice still is the difference between the overall population and a specific portfolio of a life insurer. One insightful possibility could be to further analyze the two-layer methodology for improving LC by machine learning. This could serve as a measure for portfolio modeling by first estimating the overall population drift using aggregate data, such as in this thesis. In a second step, one could estimate the portfolio drift as an improvement factor with respect to individual portfolio data. This would have the advantage that two reference levels are used, so that substructures can possibly be discovered.
Additionally, there is also a huge scope for new models if the cause-of-death information is included. Particularly with regard to COVID-19, this could provide new insights and a more detailed view of prognosis as well as a detailed capturing of excess mortality. This could be integrated relatively conveniently as an extra feature in the context of the GAM developed in this thesis and it would certainly be interesting to investigate the differences to the analyses already done. There are authors who further acknowledge that there is a close relationship between socioeconomic circumstances and causes of death. This would open another field of investigation.
There are still many unanswered vital questions to be addressed in this field of research. We encourage further research into this topic as it is crucial for insurance companies to gain further expertise on the evolution of longevity using efficient techniques. Predicting mortality is key to the longevity insurance business, since insurers in particular take on the risk for longer life. In practice, the Lee-Carter model is often used. However, as we have seen in the breakpoint analysis, the period effect is not in fact a linear drift, so a linear extrapolation of the trend could be hazardous. Therefore, it makes sense to conduct a sensitivity analysis and calculate better risk prices using accurate information, which can be improved using the methods discussed in this thesis, among others. Finally, not taking COVID-19 into account would lead to an erroneous risk assessment, which suggests the creation of different scenarios, as done in this thesis, also for later studies.

References

Antonio, K., Devriendt, S., Robben, J. and Szajder, D. (2020). Assessing the impact of COVID-19 on the iabe 2020 mortality projection: a scenario analysis, KU Leuven.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes, Econometrica.

Bai, J. and Perron, P. (2003). Computation and Analysis of Multiple Structural Change Models, howpublished of Applied Econometrics.

Barone-Adesi, F., Gasparrini, A., Vizzini, L., Merletti, F. and Richiardi, L. (2011). Effects of italian smoking regulation on rates of hospital admission for acute coronary events: A country-wide study, PLoS ONE.

Bender, A. (2018). Flexible Modeling of Time-to-event Data and Exposure-Lag-Response Associations, Ludwig-Maximilians-Universität München.

Bernstein, L. (2018). Life expectancy declines again, a dismal trend not seen since World War 1, Washington Post.

Bots, M. L. and Grobbee, D. E. (1996). Decline of coronary heart disease mortality in the Netherlands from 1978 to 1985: contribution of medical care and changes over time in presence of major cardiovascular risk factors, howpublished of cardiovascular risk.

Boudourakis, L. and Uppal, A. (2021). Decreased COVID-19 Mortality—A Cause for Optimism, JAMA Internal Medicine.

Box, G. E. P. and Jenkins, G. M. (1976). Time series analysis, control, and forecasting, San Francisco, CA: Holden Day.

Breiman, L. (2001). Random forests, Machine Learning.

Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). Classification and regression trees, Boca Raton: CRC Press.

Brouhns, N., Denuit, M. and Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables, Insurance: Mathematics and Economics.

Brown, J., Liebman, J. and Pollet, J. (2002). The Distributional Aspects of Social Security and Social Security Reform, National Bureau of Economic Research.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A. and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States, North American Actuarial howpublished.

Cairns, A. J., Blake, D. and Down, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration, howpublished of Risk and Insurance.

Carnes, B., Olshansky, S. and Grahn, D. (1996). Continuing the search for a law of mortality, Population and Development Review.

Carstensen, B. (2007). Age–period–cohort models for the Lexis diagram, Statistics in medicine.

Casella, G. and Berger, R. (2002). *Statistical Inference*, Duxbury advanced series in statistics and decision sciences, Thomson Learning.

Clayton, D. and Schifflers, E. (1987). Models for temporal variation in cancer rates. II: Age-period-cohort models, Statistics in Medicine.

Clements, M. S., Armstrong, B. K. and Moolgavkar, S. H. (2005). Lung cancer rate predictions using generalized additive models, Biostatistics.

Coelho, E. and Nunes, L. C. (2011). Forecasting mortality in the event of a structural change, howpublished of the Royal Statistical Society: Series A (Statistics in Society).

Cohu, S., Lequet-Slama, D. and Volovitch, P. (2011). The Netherlands: reform of the health system based on competition and privatisation, Revue francaise des affaires sociales.

Crimmins, E., Shim, H., Zhang, Y. and Kim, J. (2019). Differences between Men and Women in Mortality and the Health Dimensions of the Morbidity Process, Clin Chem.

Currie, I. D. (2006). *Smoothing and forecasting mortality rates with P-splines.* **URL:** *http://www.macs.hw.ac.uk/ iain/research/talks/Mortality.pdf*

Czernichow, S., Renuy, A., Rives-Lange, C., Carette, C., Airagnes, G., Wiernik, E., Ozguler, A., Kab, S., Goldberg, M., Zins, M. and Matta, J. (2021). Evolution of the prevalence of obesity in the adult population in France, Nature.

Deprez, P., Shevchenko, P. V. and Wüthrich, M. V. (2017). Machine learning techniques for mortality modeling, European Actuarial howpublished.

Dept of Transportation (US), National Highway Traffic Safety Administration (NHTSA) (2010). Traffic Tech: Primary laws and fine levels are associated with increases in seat belt use, Washington (DC): NHTSA.

Dowell, D., Arias, E., Kochanek, K., Anderson, R., Guy, Gery P., J., Losby, J. L. and Baldwin, G. (2017). Contribution of Opioid-Involved Poisoning to the Change in Life Expectancy in the United States, 2000-2015, JAMA.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties, Statist. Sci.

Emanuel, E., Gudbranson, E., Parys, J. V., Gortz, M., Helgeland, J. and Skinner, J. (2021). Comparing Health Outcomes of Privileged US Citizens With Those of Average Residents of Other Developed Countries, JAMA Intern Med.

Fahrmeir, L., Kneib, T., Lang, S. et al. (2007). Regression, Springer.

Girosi, F. and King, G. (2007). Understanding the Lee-Carter Mortality Forecasting Method.
URL: *http://gking.harvard.edu/files/abs/lc-abs.shtml*

Girosi, F. and King, G. (2008). Demographic Forecasting, Princeton University Press.

Gompertz, B. (1825). 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, Philosophical Transactions of the Royal Society of London.

Greenwell, B., Boehmke, B. and Cunningham, J. (2020). *Generalized Boosted Regression Models*. last retrieved 28.07.2021.
URL: *https://cran.r-project.org/web/packages/gbm/gbm.pdf*

Grewa, M. S. (2011). Kalman filtering, Springer Berlin Heidelberg.

Haberman, S. and Renshaw, A. (2011). A comparative study of parametric mortality projection models, Insurance: Mathematics and Economics.

Harvey, A. C. (1990). Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge university press.

Hastie, J., Hastie, T. and Tibshirani, R. (2016). The Elements of Statistical Learning, New York: Springer.

Hastie, T. and Tibshirani, R. (1987). Generalized Additive Models: Some Applications, howpublished of the American Statistical Association.

Hastie, T. and Tibshirani, R. (1990). Generalized Additive Models, Boca Raton: Routledge.

HMD (2021). Human Mortality Database, University of California Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).

Hobcraft, J., Menken, J. and Preston, S. (1982). Age, period, and cohort effects in demography: a review, Population Index.

Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates, Biometrics.

Horiuchi, S. (1983). The long-term impact of war on mortality: old-age mortality of the First World War survivors in the Federal Republic of Germany, Population Bulletin of the United Nations.

Hyndman, R. J. and Khandakar, Y. (2008). *Automatic Time Series Forecasting: The forecast Package for R*. last retrieved 28.07.2021.
URL: *https://cran.r-project.org/web/packages/forecast/vignettes/JSS2008.pdf*

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2017). An Introduction to Statistical Learning: With Applications in R, New York: Springer.

Johns Hopkins (2021). *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU).* last retrieved 30.05.2021. **URL:** *https://coronavirus.jhu.edu*

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems, Transactions of the ASME–howpublished of Basic Engineering.

Kannisto, V. (1996). The advancing frontier of survival, Odense Monographs on Population Aging. Odense University Press.

Kannisto, V., Lauritsen, J., Thatcher, A. R. and Vaupel, J. W. (1994). Reductions in mortality at advanced ages: several decades of evidence from 27 countries, Population and Development Review.

Keyes, K. and Li, G. (2019). Age–Period–Cohort Modeling, Injury Research: Springer.

Kunzler, A., Röthke, N., Günthner, L. and Stoffers-Winterling, J. (2021). Mental burden and its risk and protective factors during the early phase of the SARS-CoV-2 pandemic: systematic review and meta-analyses, Global Health.

Lee, R. D. and Carter, L. (1992). Modeling and forecasting US mortality, howpublished of the American Statistical Association.

Lee, R. and Miller, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality, Demography.

Levantesi, S. and Pizzorusso, V. (2019). Application of machine learning to mortality modeling and forecasting, Risks.

Liaw, A. and Wiener, M. (2018). *Breiman and Cutler's Random Forests for Classification and Regression.* last retrieved 28.07.2021. **URL:** *https://cran.r-project.org/web/packages/randomForest/randomForest.pdf*

Lovasz, E. (2011). Analysis of Finnish and Swedish mortality data with stochastic mortality models, European Actuarial howpublished.

Martikainen, P. and Valkonen, T. (1996). Excess mortality of unemployed men and women during a period of rapidly increasing unemployment, Lancet.

Mohan, B. and Nambiar, V. (2020). COVID-19: An Insight into SARS-CoV-2 Pandemic Originated at Wuhan City in Hubei Province of China, howpublished of Infectious Diseases and Epidemiology.

Moreno-Serra, R. and Wagsta, A. (2010). System-wide impacts of hospital payment reforms: evidence from Central and Eastern Europe and Central Asia, howpublished of Health Economics.

Moritz, S. (2021). *Time Series Missing Value Imputation.* last retrieved 28.07.2021. **URL:** *https://cran.r-project.org/web/packages/imputeTS/imputeTS.pdf*

Murphy, M. (2009). The golden generations in historical context, British Actuarial howpublished.

Murphy, M. (2010). Reexamining the dominance of birth cohort effects on mortality, Population and Development Review.

OECD (2020a). *Alcohol consumption (indicator)*. last retrieved 28.07.2021.
URL: *doi: 10.1787/e6895909-en*

OECD (2020b). *Tax revenue (indicator)*. last retrieved 28.07.2021.
URL: *doi: 10.1787/d98b8cf5-en*

OECD (2020c). *Unemployment rate (indicator)*. last retrieved 28.07.2021.
URL: *doi: 10.1787/52570002-en*

OECD (2021). *Organisation for Economic Co-operation and Development*. last retrieved 30.06.2021.
URL: *https://data.oecd.org/*

Olshansky, S. and Carnes, B. (1997). Ever since Gompertz, Demography.

Oppermann, M. (1995). Travel Life Cycle, Annals of Tourism Research.

Pascariu, M. D. (2018). Modelling and forecasting mortality, Sydddansk Universitet.

Perls, T. T. and Fretts, R. C. (1998). Why Women Live Longer than Men - What gives women the extra years?, Scientific American.

Polack, F. P., Thomas, S. J., Kitchin, N. and Absalon (2020). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine, New England howpublished of Medicine.
URL: *https://doi.org/10.1056/NEJMoa2034577*

Reither, E., Hauser, R. and Yang, Y. (2009). Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States, Social science and medicine.

Richman, R., Scognamiglio, S., Wüthrich, M. and Perla, F. (2020). Time-Series Forecasting of Mortality Rates using Deep Learning, Taylor & Francis.

Rosella, L., Calzavara, A. and Frank, J. (2016). Narrowing mortality gap between men and women over two decades: a registry-based study in Ontario, Canada BMJ Open.

Stekhoven, D. J. (2016). *Nonparametric Missing Value Imputation using Random Forest*. last retrieved 28.07.2021.
URL: *https://cran.r-project.org/web/packages/missForest/missForest.pdf*

STMF (2021). Short-term Mortality Fluctuations, University of California Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).

Strobl, C., Malley, J. D. and Tutz, G. (2009). An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests, Psychological Methods.

Sudre, C., Murray, B. and Varsavsky, T. (2021). Attributes and predictors of long COVID, Nat Med 27.

Therneau, T. M. and Atkinson, E. J. (2017). *An Introduction to Recursive Partitioning Using the RPART Routines*. last retrieved 28.07.2021.
URL: *https://cran.r-project.org/web/packages/rpart/rpart.pdf*

UIS (2021). *UNESCO Institute for Statistics (UIS)*. last retrieved 30.06.2021.
URL: *uis.unesco.org*

UNDP (2021). *Human Development Reports*. last retrieved 30.06.2021.
URL: *http://hdr.undp.org/*

van Berkum, F., Antonio, K. and Vellekoop, M. (2013). Structural changes in mortality rates with an application to Dutch and Belgian data, SSRN Electronic howpublished.

Vaupel, J. W., Carey, J. R., Christensen, K. and Johnson, T. E. (1998). Biodemographic trajectories of longevity, Science 280.

Villegas, A. M. and Haberman, S. (2014). On the modeling and forecasting of socioeconomic mortality differentials: an application to deprivation and mortality in England, North American Actuarial howpublished.

Villegas, M. A. R. (2015). Mortality: Modelling, Socio-Economic Differences and Basis Risk, City, University of London Institutional Repository.

Villegas, M. A. R., Millossovich, P. and Kaishev, V. (2016). StMoMo: an R package for stochastic mortality modelling. last retrieved 28.07.2021.
URL: *https://cran.r-project.org/web/packages/StMoMo/StMoMo.pdf*

Wegman, M. E. (2001). Infant Mortality in the 20th Century, Dramatic but Uneven Progress, The howpublished of Nutrition.

Weigert, M., Bauer, A., Gernert, J., Karl, M., Nalmpatian, A., Küchenhoff, H. and Schmude, J. (2021). Semiparametric APC analysis of destination choice patterns: Using generalized additive models to quantify the impact of age, period, and cohort on travel distances, Tourism Economics.

Welch, G. and Bishop, G. (2006). An Introduction to the Kalman Filter, University of North Carolina at Chapel Hill.

WHO (2020). *Prevalence of overweight among adults (indicator)*. last retrieved 28.07.2021.
URL: *doi: http://dx.doi.org/10.1016/S0140-6736(17)32129-3*

WHO (2021). *World Health Organization*. last retrieved 30.06.2021.
URL: *https://www.who.int/*

Willets, R. C. (2004). The cohort effect: insights and explanations, British Actuarial howpublished.

Wilmoth, J. R. and Horiuchi, S. (1990). Rectangularization revisited: Variability of age at death within human populations, Demography 36.

Wilson, N., Clement, C., Summers, J. A., Bannister, J. and Harper, G. (2014). Mortality of first world war military personnel: comparison of two military cohorts, BMJ Publishing Group Ltd.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, howpublished of the Royal Statistical Society. Series B (Statistical Methodology).

Wood, S. N. (2017). Generalized additive models: an introduction with R, CRC press.

Wood, S. N. (2021). *Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*. last retrieved 28.07.2021.
URL: *https://cran.r-project.org/web/packages/mgcv/mgcv.pdf*

Woolf, S. and Schoomaker, H. (2019). Life expectancy and mortality rates in the United States, JAMA.

World Bank (2021). *The World Bank Group*. last retrieved 30.04.2021.
URL: *databank.worldbank.org*

World Health Organization (2018). Global status report on alcohol and health 2018, WHO.

Yang, Y. and Land, K. C. (2013). Age-period-cohort analysis: New models, methods, and empirical applications, CRC press.

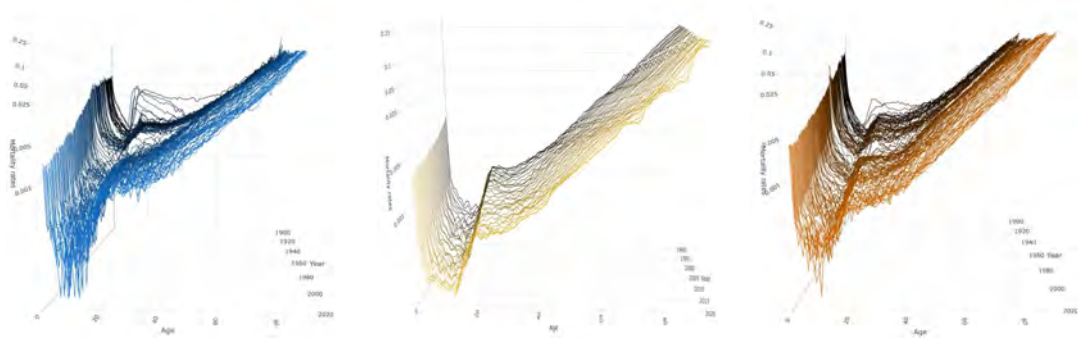
Yao, Y.-C. (1988). Estimating the number of change-points via Schwarz' criterion, Statistics & Probability Letters.

Zeileis, A. (2005). A Unified Approach to Structural Change Tests Based on ML Scores, F Statistics and OLS Residuals, Taylor & Francis.

Zeileis, A., Kleiber, C., Kramer, W. and Hornik, K. (2003). Testing and Dating of Structural Changes in Practice, Computational Statistics and Data Analysis, Computational Statistics & Data Analysis.

Zeileis, A., Leisch, F., Hornik, K. and Kleiber, C. (2002). *strucchange: An R Package for Testing for Structural Change in Linear Regression Models*. last retrieved 28.07.2021.
URL: *https://cran.r-project.org/web/packages/strucchange/vignettes/strucchange-intro.pdf*

Appendix



(a) Finnish (b) German (c) Dutch

Figure 41: Evolution of mortality rates for males aged 0–90 for the years 1900–2020 (1990–2020 in case of Germany)

Country	Female					Male				
	Class.	APC	GAM	Covariates	Continentwise	Class.	APC	GAM	Covariates	Continentwise
FIN	0.0015	0.0011	0.0012	0.0013		0.0027	0.0013	0.0013	0.0015	
DE	0.0022	0.001	0.001	0.0014		0.0021	0.001	0.001	0.0013	
ITA	0.0021	0.001	0.0012	0.0011		0.0025	0.001	0.0009	0.0011	
NLD	0.0032	0.001	0.0011	0.0011		0.0015	0.001	0.0008	0.0012	
US	0.0033	0.0013	0.0015	0.0016		0.0028	0.0011	0.0012	0.001	

Table 13: In-sample RMSE for all countries, genders and models APC as well as the GAM-based improvements with Age-Period tensor product splines, the integrated socio-economic and health-risk variables and the continentwise modeling. The fitted period is 1990-2015 for Finland, Germany, Italy, the Netherlands and the US.

Age	2019	Scenario I		Scenario II		Difference		Drop in %	
		2022	2023	2022	2023	2022	2023	2022	2023
50	29.77	29.98	30.04	29.68	29.73	0.30	0.31	1.00	1.03
51	28.88	29.09	29.14	28.79	28.84	0.30	0.30	1.03	1.03
52	27.98	28.20	28.25	27.91	27.95	0.29	0.30	1.03	1.06
53	27.10	27.31	27.36	27.03	27.07	0.28	0.29	1.03	1.06
54	26.22	26.44	26.49	26.16	26.20	0.28	0.29	1.06	1.09
55	25.36	25.57	25.62	25.30	25.34	0.27	0.28	1.06	1.09
56	24.49	24.71	24.75	24.44	24.48	0.27	0.27	1.09	1.09
57	23.64	23.85	23.90	23.59	23.63	0.26	0.27	1.09	1.13
58	22.79	23.01	23.05	22.76	22.79	0.25	0.26	1.09	1.13
59	21.96	22.17	22.21	21.93	21.96	0.24	0.25	1.08	1.13
60	21.13	21.34	21.38	21.11	21.14	0.23	0.24	1.08	1.12
61	20.31	20.52	20.55	20.29	20.32	0.23	0.23	1.12	1.12
62	19.50	19.71	19.74	19.49	19.52	0.22	0.22	1.12	1.11
63	18.71	18.91	18.94	18.70	18.72	0.21	0.22	1.11	1.16
64	17.92	18.12	18.14	17.91	17.94	0.21	0.20	1.16	1.10
65	17.15	17.34	17.36	17.14	17.16	0.20	0.20	1.15	1.15
66	16.41	16.57	16.59	16.38	16.40	0.19	0.19	1.15	1.15
67	15.68	15.81	15.83	15.62	15.64	0.19	0.19	1.20	1.20
68	14.95	15.07	15.08	14.88	14.90	0.19	0.18	1.26	1.19
69	14.21	14.33	14.34	14.15	14.16	0.18	0.18	1.26	1.26
70	13.46	13.61	13.62	13.43	13.44	0.18	0.18	1.32	1.32
71	12.69	12.90	12.91	12.73	12.73	0.17	0.18	1.32	1.39
72	11.95	12.20	12.21	12.03	12.04	0.17	0.17	1.39	1.39
73	11.26	11.52	11.53	11.35	11.35	0.17	0.18	1.48	1.56
74	10.69	10.85	10.86	10.67	10.68	0.18	0.18	1.66	1.66
75	10.08	10.19	10.20	10.01	10.02	0.18	0.18	1.77	1.76
76	9.49	9.54	9.55	9.37	9.37	0.17	0.18	1.78	1.88
77	8.96	8.91	8.92	8.73	8.73	0.18	0.19	2.02	2.13
78	8.40	8.28	8.30	8.11	8.11	0.17	0.19	2.05	2.29
79	7.77	7.67	7.68	7.49	7.49	0.18	0.19	2.35	2.47
80	7.12	7.06	7.08	6.89	6.89	0.17	0.19	2.41	2.68
81	6.48	6.46	6.48	6.29	6.30	0.17	0.18	2.63	2.78
82	5.85	5.87	5.89	5.70	5.71	0.17	0.18	2.90	3.06
83	5.21	5.28	5.30	5.12	5.12	0.16	0.18	3.03	3.40
84	4.52	4.68	4.70	4.54	4.54	0.14	0.16	2.99	3.40
85	3.81	4.07	4.10	3.95	3.95	0.12	0.15	2.95	3.66
86	3.20	3.45	3.47	3.34	3.35	0.11	0.12	3.19	3.46
87	2.65	2.80	2.82	2.72	2.72	0.08	0.10	2.86	3.55
88	2.04	2.11	2.12	2.05	2.05	0.06	0.07	2.84	3.30
89	1.33	1.34	1.35	1.32	1.32	0.02	0.03	1.49	2.22
90	0.49	0.49	0.49	0.49	0.49	0.00	0.00	0.00	0.00

Table 14: Life expectancy forecasts for German male population

Age	2019	Scenario I		Scenario II		Difference		Drop in %	
		2022	2023	2022	2023	2022	2023	2022	2023
50	33.10	32.81	32.79	31.83	31.81	0.98	0.98	3.00	3.00
51	32.17	31.88	31.86	30.92	30.90	0.96	0.96	3.00	3.00
52	31.23	30.96	30.93	30.96	30.00	0.00	0.93	0.00	3.00
53	30.30	30.04	30.01	30.03	29.11	0.01	0.90	0.03	3.00
54	29.37	29.12	29.09	29.11	28.22	0.01	0.87	0.03	3.00
55	28.46	28.21	28.18	28.20	27.33	0.01	0.85	0.04	3.00
56	27.54	27.31	27.27	27.29	26.45	0.02	0.82	0.07	3.00
57	26.63	26.40	26.37	26.39	25.58	0.01	0.79	0.04	3.00
58	25.72	25.51	25.48	25.49	25.48	0.02	0.00	0.08	0.00
59	24.82	24.62	24.59	24.60	24.58	0.02	0.01	0.08	0.04
60	23.93	23.74	23.70	23.71	23.70	0.03	0.00	0.13	0.00
61	23.04	22.86	22.83	22.83	22.82	0.03	0.01	0.13	0.04
62	22.16	21.99	21.95	21.96	21.94	0.03	0.01	0.14	0.05
63	21.29	21.13	21.09	21.09	21.07	0.04	0.02	0.19	0.09
64	20.42	20.27	20.23	20.23	20.21	0.04	0.02	0.20	0.10
65	19.57	19.42	19.38	19.38	19.36	0.04	0.02	0.21	0.10
66	18.73	18.58	18.54	18.53	18.51	0.05	0.03	0.27	0.16
67	17.90	17.74	17.70	17.69	17.67	0.05	0.03	0.28	0.17
68	17.06	16.91	16.87	16.86	16.83	0.05	0.04	0.30	0.24
69	16.22	16.09	16.05	16.03	16.00	0.06	0.05	0.37	0.31
70	15.38	15.28	15.24	15.22	15.19	0.06	0.05	0.39	0.33
71	14.53	14.47	14.44	14.41	14.38	0.06	0.06	0.41	0.42
72	13.69	13.68	13.64	13.61	13.57	0.07	0.07	0.51	0.51
73	12.89	12.89	12.86	12.82	12.78	0.07	0.08	0.54	0.62
74	12.15	12.12	12.08	12.03	12.00	0.09	0.08	0.74	0.66
75	11.40	11.35	11.31	11.26	11.22	0.09	0.09	0.79	0.80
76	10.70	10.59	10.56	10.50	10.46	0.09	0.10	0.85	0.95
77	10.03	9.84	9.81	9.75	9.71	0.09	0.10	0.91	1.02
78	9.35	9.11	9.08	9.01	8.96	0.10	0.12	1.10	1.32
79	8.62	8.38	8.35	8.28	8.23	0.10	0.12	1.19	1.44
80	7.87	7.66	7.64	7.56	7.52	0.10	0.12	1.31	1.57
81	7.12	6.96	6.93	6.85	6.81	0.11	0.12	1.58	1.73
82	6.39	6.26	6.24	6.16	6.11	0.10	0.13	1.60	2.08
83	5.64	5.57	5.55	5.47	5.43	0.10	0.12	1.80	2.16
84	4.86	4.89	4.87	4.79	4.76	0.10	0.11	2.04	2.26
85	4.07	4.21	4.19	4.12	4.09	0.09	0.10	2.14	2.39
86	3.40	3.53	3.51	3.45	3.42	0.08	0.09	2.27	2.56
87	2.78	2.83	2.82	2.77	2.75	0.06	0.07	2.12	2.48
88	2.12	2.11	2.10	2.07	2.05	0.04	0.05	1.90	2.38
89	1.36	1.34	1.34	1.32	1.31	0.02	0.03	1.49	2.24
90	0.49	0.49	0.49	0.49	0.48	0.00	0.01	0.00	2.04

Table 15: Life expectancy forecasts for German female population

Age	2019	Scenario I		Scenario II		Difference		Drop in %	
		2022	2023	2022	2023	2022	2023	2022	2023
50	31.25	31.58	31.72	31.13	31.31	0.45	0.41	1.42	1.29
51	30.32	30.66	30.79	30.20	30.38	0.46	0.41	1.50	1.33
52	29.40	29.74	29.87	29.28	29.45	0.46	0.42	1.55	1.41
53	28.48	28.82	28.95	28.36	28.53	0.46	0.42	1.60	1.45
54	27.57	27.91	28.04	27.45	27.61	0.46	0.43	1.65	1.53
55	26.65	27.01	27.14	26.54	26.70	0.47	0.44	1.74	1.62
56	25.76	26.11	26.23	25.64	25.80	0.47	0.43	1.80	1.64
57	24.86	25.21	25.34	24.74	24.89	0.47	0.45	1.86	1.78
58	23.99	24.33	24.45	23.85	24.00	0.48	0.45	1.97	1.84
59	23.11	23.44	23.56	22.96	23.11	0.48	0.45	2.05	1.91
60	22.25	22.57	22.69	22.08	22.23	0.49	0.46	2.17	2.03
61	21.39	21.70	21.82	21.21	21.35	0.49	0.47	2.26	2.15
62	20.54	20.84	20.95	20.35	20.48	0.49	0.47	2.35	2.24
63	19.70	19.99	20.10	19.50	19.62	0.49	0.48	2.45	2.39
64	18.86	19.14	19.25	18.65	18.77	0.49	0.48	2.56	2.49
65	18.05	18.31	18.41	17.81	17.93	0.50	0.48	2.73	2.61
66	17.25	17.48	17.58	16.98	17.09	0.50	0.49	2.86	2.79
67	16.45	16.67	16.76	16.17	16.27	0.50	0.49	3.00	2.92
68	15.65	15.86	15.95	15.36	15.45	0.50	0.50	3.15	3.13
69	14.87	15.06	15.15	14.57	14.65	0.49	0.50	3.25	3.30
70	14.10	14.28	14.36	13.79	13.86	0.49	0.50	3.43	3.48
71	13.35	13.51	13.58	13.02	13.09	0.49	0.49	3.63	3.61
72	12.59	12.74	12.82	12.26	12.33	0.48	0.49	3.77	3.82
73	11.85	12.00	12.06	11.52	11.58	0.48	0.48	4.00	3.98
74	11.12	11.26	11.32	10.79	10.84	0.47	0.48	4.17	4.24
75	10.41	10.54	10.60	10.08	10.13	0.46	0.47	4.36	4.43
76	9.73	9.83	9.88	9.39	9.42	0.44	0.46	4.48	4.66
77	9.04	9.13	9.18	8.71	8.74	0.42	0.44	4.60	4.79
78	8.37	8.45	8.50	8.05	8.07	0.40	0.43	4.73	5.06
79	7.70	7.79	7.82	7.40	7.42	0.39	0.40	5.01	5.12
80	7.06	7.13	7.16	6.77	6.78	0.36	0.38	5.05	5.31
81	6.44	6.49	6.52	6.16	6.16	0.33	0.36	5.08	5.52
82	5.81	5.86	5.88	5.56	5.56	0.30	0.32	5.12	5.44
83	5.19	5.24	5.25	4.97	4.96	0.27	0.29	5.15	5.52
84	4.58	4.62	4.63	4.39	4.38	0.23	0.25	4.98	5.40
85	3.97	4.00	4.01	3.81	3.80	0.19	0.21	4.75	5.24
86	3.37	3.38	3.38	3.22	3.22	0.16	0.16	4.73	4.73
87	2.72	2.73	2.74	2.62	2.62	0.11	0.12	4.03	4.38
88	2.06	2.06	2.06	1.99	1.98	0.07	0.08	3.40	3.88
89	1.33	1.32	1.32	1.29	1.29	0.03	0.03	2.27	2.27
90	0.49	0.49	0.49	0.48	0.48	0.01	0.01	2.04	2.04

Table 16: Life expectancy forecasts for Dutch male population

Age	2019	Scenario I		Scenario II		Difference		Drop in %	
		2022	2023	2022	2023	2022	2023	2022	2023
50	33.36	33.30	33.39	32.30	32.39	1.00	1.00	3.00	3.00
51	32.41	32.37	32.46	31.40	31.49	0.97	0.97	3.00	3.00
52	31.47	31.44	31.53	30.50	30.58	0.94	0.95	3.00	3.00
53	30.55	30.52	30.60	29.60	29.68	0.92	0.92	3.00	3.00
54	29.62	29.60	29.68	29.60	28.79	0.00	0.89	0.00	3.00
55	28.69	28.69	28.76	28.67	27.90	0.02	0.86	0.07	3.00
56	27.78	27.77	27.85	27.75	27.01	0.02	0.84	0.07	3.00
57	26.86	26.87	26.94	26.83	26.13	0.04	0.81	0.15	3.00
58	25.95	25.97	26.04	25.91	25.26	0.06	0.78	0.23	3.00
59	25.06	25.07	25.14	25.00	24.39	0.07	0.75	0.28	3.00
60	24.17	24.18	24.24	24.09	24.24	0.09	0.00	0.37	0.00
61	23.28	23.29	23.35	23.19	23.33	0.10	0.02	0.43	0.09
62	22.41	22.41	22.47	22.29	22.43	0.12	0.04	0.54	0.18
63	21.54	21.53	21.59	21.40	21.53	0.13	0.06	0.60	0.28
64	20.68	20.66	20.72	20.52	20.64	0.14	0.08	0.68	0.39
65	19.82	19.80	19.85	19.64	19.75	0.16	0.10	0.81	0.50
66	18.98	18.94	18.99	18.77	18.87	0.17	0.12	0.90	0.63
67	18.13	18.10	18.14	17.90	18.00	0.20	0.14	1.10	0.77
68	17.29	17.25	17.30	17.05	17.14	0.20	0.16	1.16	0.92
69	16.47	16.42	16.46	16.20	16.28	0.22	0.18	1.34	1.09
70	15.64	15.59	15.63	15.36	15.43	0.23	0.20	1.48	1.28
71	14.83	14.78	14.81	14.54	14.60	0.24	0.21	1.62	1.42
72	14.02	13.97	14.00	13.72	13.77	0.25	0.23	1.79	1.64
73	13.22	13.17	13.20	12.91	12.95	0.26	0.25	1.97	1.89
74	12.43	12.38	12.40	12.11	12.15	0.27	0.25	2.18	2.02
75	11.64	11.60	11.62	11.33	11.36	0.27	0.26	2.33	2.24
76	10.88	10.83	10.85	10.55	10.58	0.28	0.27	2.59	2.49
77	10.12	10.07	10.09	9.79	9.81	0.28	0.28	2.78	2.78
78	9.36	9.32	9.33	9.05	9.06	0.27	0.27	2.90	2.89
79	8.62	8.58	8.59	8.31	8.32	0.27	0.27	3.15	3.14
80	7.88	7.85	7.86	7.59	7.60	0.26	0.26	3.31	3.31
81	7.15	7.13	7.14	6.89	6.89	0.24	0.25	3.37	3.50
82	6.44	6.42	6.43	6.19	6.19	0.23	0.24	3.58	3.73
83	5.74	5.71	5.72	5.51	5.50	0.20	0.22	3.50	3.85
84	5.04	5.01	5.02	4.83	4.83	0.18	0.19	3.59	3.78
85	4.34	4.32	4.32	4.16	4.15	0.16	0.17	3.70	3.94
86	3.63	3.61	3.61	3.49	3.48	0.12	0.13	3.32	3.60
87	2.91	2.89	2.89	2.80	2.80	0.09	0.09	3.11	3.11
88	2.16	2.15	2.15	2.09	2.09	0.06	0.06	2.79	2.79
89	1.37	1.36	1.36	1.33	1.33	0.03	0.03	2.21	2.21
90	0.49	0.49	0.49	0.49	0.49	0.00	0.00	0.00	0.00

Table 17: Life expectancy forecasts for Dutch female population

Age	2019	Scenario I		Scenario II		Difference		Drop in %	
		2022	2023	2022	2023	2022	2023	2022	2023
50	31.54	31.97	32.09	30.81	30.93	1.16	1.16	3.63	3.61
51	30.62	31.05	31.17	29.89	30.01	1.16	1.16	3.74	3.72
52	29.70	30.13	30.25	28.97	29.09	1.16	1.16	3.85	3.83
53	28.79	29.21	29.33	28.06	28.17	1.15	1.16	3.94	3.95
54	27.87	28.30	28.42	27.15	27.26	1.15	1.16	4.06	4.08
55	26.96	27.40	27.51	26.25	26.35	1.15	1.16	4.20	4.22
56	26.06	26.50	26.61	25.36	25.45	1.14	1.16	4.30	4.36
57	25.16	25.60	25.71	24.47	24.56	1.13	1.15	4.41	4.47
58	24.28	24.71	24.82	23.58	23.67	1.13	1.15	4.57	4.63
59	23.40	23.83	23.93	22.71	22.79	1.12	1.14	4.70	4.76
60	22.53	22.95	23.05	21.84	21.92	1.11	1.13	4.84	4.90
61	21.67	22.08	22.18	20.98	21.05	1.10	1.13	4.98	5.09
62	20.81	21.21	21.31	20.12	20.19	1.09	1.12	5.14	5.26
63	19.97	20.36	20.45	19.28	19.35	1.08	1.10	5.30	5.38
64	19.13	19.51	19.60	18.45	18.51	1.06	1.09	5.43	5.56
65	18.31	18.67	18.76	17.62	17.68	1.05	1.08	5.62	5.76
66	17.51	17.84	17.92	16.81	16.86	1.03	1.06	5.77	5.92
67	16.72	17.02	17.10	16.01	16.06	1.01	1.04	5.93	6.08
68	15.96	16.20	16.28	15.22	15.26	0.98	1.02	6.05	6.27
69	15.18	15.40	15.48	14.44	14.48	0.96	1.00	6.23	6.46
70	14.40	14.61	14.68	13.67	13.71	0.94	0.97	6.43	6.61
71	13.61	13.83	13.90	12.92	12.96	0.91	0.94	6.58	6.76
72	12.82	13.07	13.13	12.19	12.22	0.88	0.91	6.73	6.93
73	12.04	12.31	12.37	11.46	11.49	0.85	0.88	6.90	7.11
74	11.33	11.57	11.62	10.75	10.78	0.82	0.84	7.09	7.23
75	10.63	10.84	10.88	10.06	10.08	0.78	0.80	7.20	7.35
76	9.96	10.12	10.16	9.38	9.40	0.74	0.76	7.31	7.48
77	9.31	9.41	9.46	8.72	8.74	0.69	0.72	7.33	7.61
78	8.65	8.72	8.76	8.07	8.09	0.65	0.67	7.45	7.65
79	7.97	8.04	8.08	7.44	7.46	0.60	0.62	7.46	7.67
80	7.28	7.37	7.41	6.83	6.84	0.54	0.57	7.33	7.69
81	6.60	6.72	6.75	6.22	6.23	0.50	0.52	7.44	7.70
82	5.94	6.07	6.09	5.63	5.64	0.44	0.45	7.25	7.39
83	5.31	5.43	5.45	5.04	5.05	0.39	0.40	7.18	7.34
84	4.66	4.79	4.81	4.46	4.47	0.33	0.34	6.89	7.07
85	4.00	4.14	4.16	3.88	3.89	0.26	0.27	6.28	6.49
86	3.40	3.49	3.50	3.29	3.30	0.20	0.20	5.73	5.71
87	2.78	2.82	2.83	2.68	2.68	0.14	0.15	4.96	5.30
88	2.10	2.11	2.12	2.03	2.03	0.08	0.09	3.79	4.25
89	1.34	1.34	1.35	1.31	1.31	0.03	0.04	2.24	2.96
90	0.49	0.49	0.49	0.48	0.48	0.01	0.01	2.04	2.04

Table 18: Life expectancy forecasts for Italian male population

Age	2019	Scenario I		Scenario II		Difference		Drop in %	
		2022	2023	2022	2023	2022	2023	2022	2023
50	34.41	34.30	34.36	33.88	34.00	0.42	0.36	1.22	1.05
51	33.46	33.37	33.42	32.94	33.05	0.43	0.37	1.29	1.11
52	32.51	32.43	32.49	31.99	32.10	0.44	0.39	1.36	1.20
53	31.56	31.50	31.55	31.05	31.16	0.45	0.39	1.43	1.24
54	30.61	30.57	30.62	30.11	30.21	0.46	0.41	1.50	1.34
55	29.67	29.64	29.69	29.17	29.27	0.47	0.42	1.59	1.41
56	28.73	28.71	28.76	28.24	28.34	0.47	0.42	1.64	1.46
57	27.80	27.79	27.84	27.30	27.40	0.49	0.44	1.76	1.58
58	26.87	26.87	26.91	26.38	26.47	0.49	0.44	1.82	1.64
59	25.95	25.95	26.00	25.45	25.54	0.50	0.46	1.93	1.77
60	25.03	25.04	25.08	24.53	24.62	0.51	0.46	2.04	1.83
61	24.12	24.13	24.17	23.62	23.70	0.51	0.47	2.11	1.94
62	23.21	23.22	23.26	22.71	22.78	0.51	0.48	2.20	2.06
63	22.31	22.32	22.36	21.80	21.87	0.52	0.49	2.33	2.19
64	21.41	21.42	21.46	20.90	20.96	0.52	0.50	2.43	2.33
65	20.53	20.53	20.56	20.00	20.06	0.53	0.50	2.58	2.43
66	19.65	19.64	19.67	19.11	19.17	0.53	0.50	2.70	2.54
67	18.78	18.76	18.79	18.23	18.28	0.53	0.51	2.83	2.71
68	17.92	17.89	17.91	17.36	17.40	0.53	0.51	2.96	2.85
69	17.05	17.02	17.04	16.49	16.53	0.53	0.51	3.11	2.99
70	16.19	16.16	16.18	15.63	15.66	0.53	0.52	3.28	3.21
71	15.32	15.30	15.32	14.78	14.81	0.52	0.51	3.40	3.33
72	14.45	14.46	14.47	13.94	13.96	0.52	0.51	3.60	3.52
73	13.58	13.62	13.63	13.11	13.12	0.51	0.51	3.74	3.74
74	12.77	12.79	12.79	12.29	12.30	0.50	0.49	3.91	3.83
75	11.96	11.97	11.97	11.49	11.49	0.48	0.48	4.01	4.01
76	11.18	11.16	11.16	10.69	10.69	0.47	0.47	4.21	4.21
77	10.42	10.36	10.36	9.91	9.90	0.45	0.46	4.34	4.44
78	9.65	9.57	9.57	9.14	9.13	0.43	0.44	4.49	4.60
79	8.86	8.79	8.79	8.39	8.37	0.40	0.42	4.55	4.78
80	8.08	8.03	8.02	7.65	7.63	0.38	0.39	4.73	4.86
81	7.31	7.28	7.26	6.92	6.90	0.36	0.36	4.95	4.96
82	6.54	6.53	6.52	6.21	6.19	0.32	0.33	4.90	5.06
83	5.79	5.80	5.79	5.51	5.49	0.29	0.30	5.00	5.18
84	5.04	5.07	5.06	4.83	4.80	0.24	0.26	4.73	5.14
85	4.27	4.35	4.34	4.15	4.12	0.20	0.22	4.60	5.07
86	3.60	3.63	3.62	3.47	3.45	0.16	0.17	4.41	4.70
87	2.91	2.90	2.89	2.79	2.77	0.11	0.12	3.79	4.15
88	2.18	2.15	2.14	2.08	2.07	0.07	0.07	3.26	3.27
89	1.37	1.36	1.35	1.33	1.32	0.03	0.03	2.21	2.22
90	0.49	0.49	0.49	0.49	0.49	0.00	0.00	0.00	0.00

Table 19: Life expectancy forecasts for Italian female population

Age	2019	Scenario I		Scenario II		Difference		Drop in %	
		2022	2023	2022	2023	2022	2023	2022	2023
50	30.55	31.11	31.28	30.82	31.02	0.29	0.26	0.93	0.83
51	29.64	30.21	30.38	29.92	30.12	0.29	0.26	0.96	0.86
52	28.73	29.31	29.48	29.03	29.23	0.28	0.25	0.96	0.85
53	27.84	28.42	28.58	28.15	28.35	0.27	0.23	0.95	0.80
54	26.94	27.53	27.69	27.27	27.47	0.26	0.22	0.94	0.79
55	26.06	26.65	26.81	26.40	26.60	0.25	0.21	0.94	0.78
56	25.20	25.77	25.93	25.54	25.74	0.23	0.19	0.89	0.73
57	24.34	24.90	25.06	24.69	24.89	0.21	0.17	0.84	0.68
58	23.49	24.04	24.20	23.84	24.05	0.20	0.15	0.83	0.62
59	22.64	23.19	23.34	23.00	23.21	0.19	0.13	0.82	0.56
60	21.80	22.34	22.49	22.18	22.38	0.16	0.11	0.72	0.49
61	20.96	21.50	21.65	21.36	21.57	0.14	0.08	0.65	0.37
62	20.13	20.67	20.82	20.55	20.76	0.12	0.06	0.58	0.29
63	19.32	19.85	19.99	19.74	19.96	0.11	0.03	0.55	0.15
64	18.52	19.03	19.18	18.95	19.16	0.08	0.02	0.42	0.10
65	17.74	18.23	18.37	18.17	17.82	0.06	0.55	0.33	3.00
66	16.96	17.43	17.57	17.39	17.04	0.04	0.53	0.23	3.00
67	16.20	16.65	16.78	16.63	16.28	0.02	0.50	0.12	3.00
68	15.44	15.87	16.00	15.87	15.52	0.00	0.48	0.00	3.00
69	14.69	15.10	15.23	14.65	14.77	0.45	0.46	3.00	3.00
70	13.96	14.35	14.47	13.92	14.04	0.43	0.43	3.00	3.00
71	13.26	13.60	13.73	13.19	13.32	0.41	0.41	3.00	3.00
72	12.55	12.87	12.99	12.48	12.60	0.39	0.39	3.00	3.00
73	11.84	12.15	12.26	11.79	11.89	0.36	0.37	3.00	3.00
74	11.14	11.44	11.55	11.10	11.20	0.34	0.35	3.00	3.00
75	10.45	10.74	10.84	10.42	10.51	0.32	0.33	3.00	3.00
76	9.78	10.05	10.15	9.75	9.85	0.30	0.30	3.00	3.00
77	9.12	9.38	9.47	9.10	9.19	0.28	0.28	3.00	3.00
78	8.47	8.71	8.80	8.45	8.54	0.26	0.26	3.00	3.00
79	7.82	8.05	8.14	7.81	7.90	0.24	0.24	3.00	3.00
80	7.17	7.40	7.48	7.18	7.26	0.22	0.22	3.00	3.00
81	6.55	6.76	6.83	6.56	6.63	0.20	0.20	3.00	3.00
82	5.92	6.12	6.19	5.94	6.00	0.18	0.19	3.00	3.00
83	5.31	5.49	5.55	5.33	5.38	0.16	0.17	3.00	3.00
84	4.68	4.85	4.90	4.85	4.75	0.00	0.15	0.00	3.00
85	4.06	4.21	4.25	4.20	4.25	0.01	0.00	0.24	0.00
86	3.42	3.55	3.58	3.53	3.58	0.02	0.00	0.56	0.00
87	2.78	2.86	2.89	2.85	2.88	0.01	0.01	0.35	0.35
88	2.09	2.14	2.15	2.13	2.14	0.01	0.01	0.47	0.47
89	1.33	1.36	1.36	1.35	1.36	0.01	0.00	0.74	0.00
90	0.49	0.49	0.49	0.49	0.49	0.00	0.00	0.00	0.00

Table 20: Life expectancy forecasts for Finnish male population

Age	2019	Scenario I		Scenario II		Difference		Drop in %	
		2022	2023	2022	2023	2022	2023	2022	2023
50	34.13	34.19	34.32	34.03	34.18	0.16	0.14	0.47	0.41
51	33.19	33.26	33.39	33.10	33.24	0.16	0.15	0.48	0.45
52	32.24	32.33	32.46	32.16	32.31	0.17	0.15	0.53	0.46
53	31.30	31.41	31.54	31.23	31.38	0.18	0.16	0.57	0.51
54	30.39	30.49	30.61	30.31	30.45	0.18	0.16	0.59	0.52
55	29.45	29.57	29.70	29.39	29.53	0.18	0.17	0.61	0.57
56	28.51	28.66	28.78	28.47	28.61	0.19	0.17	0.66	0.59
57	27.59	27.75	27.87	27.55	27.69	0.20	0.18	0.72	0.65
58	26.66	26.84	26.96	26.64	26.78	0.20	0.18	0.75	0.67
59	25.74	25.94	26.06	25.74	25.87	0.20	0.19	0.77	0.73
60	24.84	25.04	25.16	24.83	24.97	0.21	0.19	0.84	0.76
61	23.94	24.14	24.27	23.94	24.07	0.20	0.20	0.83	0.82
62	23.05	23.25	23.38	23.04	23.18	0.21	0.20	0.90	0.86
63	22.14	22.37	22.49	22.16	22.29	0.21	0.20	0.94	0.89
64	21.25	21.49	21.61	21.27	21.41	0.22	0.20	1.02	0.93
65	20.37	20.61	20.73	20.40	20.53	0.21	0.20	1.02	0.96
66	19.52	19.74	19.86	19.53	19.65	0.21	0.21	1.06	1.06
67	18.67	18.88	18.99	18.66	18.79	0.22	0.20	1.17	1.05
68	17.80	18.02	18.13	17.80	17.93	0.22	0.20	1.22	1.10
69	16.95	17.17	17.28	16.95	17.07	0.22	0.21	1.28	1.22
70	16.10	16.32	16.43	16.11	16.22	0.21	0.21	1.29	1.28
71	15.25	15.48	15.59	15.27	15.38	0.21	0.21	1.36	1.35
72	14.43	14.65	14.75	14.44	14.55	0.21	0.20	1.43	1.36
73	13.60	13.82	13.92	13.62	13.72	0.20	0.20	1.45	1.44
74	12.79	13.00	13.10	12.80	12.90	0.20	0.20	1.54	1.53
75	11.97	12.19	12.29	12.00	12.09	0.19	0.20	1.56	1.63
76	11.18	11.39	11.48	11.20	11.29	0.19	0.19	1.67	1.66
77	10.39	10.60	10.68	10.41	10.50	0.19	0.18	1.79	1.69
78	9.62	9.81	9.89	9.64	9.71	0.17	0.18	1.73	1.82
79	8.83	9.03	9.11	8.87	8.94	0.16	0.17	1.77	1.87
80	8.09	8.27	8.33	8.11	8.17	0.16	0.16	1.93	1.92
81	7.35	7.50	7.56	7.36	7.42	0.14	0.14	1.87	1.85
82	6.61	6.75	6.80	6.62	6.67	0.13	0.13	1.93	1.91
83	5.87	6.00	6.05	5.88	5.93	0.12	0.12	2.00	1.98
84	5.16	5.25	5.29	5.15	5.19	0.10	0.10	1.90	1.89
85	4.42	4.51	4.54	4.42	4.45	0.09	0.09	2.00	1.98
86	3.69	3.76	3.78	3.69	3.71	0.07	0.07	1.86	1.85
87	2.95	3.00	3.01	2.95	2.96	0.05	0.05	1.67	1.66
88	2.19	2.21	2.22	2.18	2.19	0.03	0.03	1.36	1.35
89	1.38	1.38	1.39	1.37	1.37	0.01	0.02	0.72	1.44
90	0.49	0.49	0.49	0.49	0.49	0.00	0.00	0.00	0.00

Table 21: Life expectancy forecasts for Finnish female population

List of Abbreviations

AIC	Akaike Information Criterion
APC	Age-Period-Cohort
ARIMA	Autoregressive Integrated Moving Average
BIC	Bayesian Information Criterion
BMI	Body Mass Index
CART	Classification and Regression Trees
CBD	Cairns-Blake-Dowd
GAM	Generalized Additive Model
GAPC	Generalized Age-Period-Cohort
GCV	Generalized Cross Validation
GDP	Gross Domestic Product
GLM	Generalized Linear Model
HMD	Human Mortality Database
i.i.d.	independent and identically distributed
LC	Lee Carter
M7	Quadratic CBD with cohort effects
MAPE	Mean Absolute Percentage Error
OECD	Organisation for Economic Co-operation and Development
P-IRLS	Penalized Iteratively Re-weighted Least Squares
REML	Restricted Maximum Likelihood
RMSE	Root Mean Square Error
STMF	Short Term Mortality Fluctuations
UIS	UNESCO Institute of Statistics
WHO	World Health Organization

List of Figures

List of Tables

Declaration of Authorship

I hereby confirm that I have written the accompanying thesis by myself, without contributions from any sources other than those cited in the text and acknowledgements. This also applies to all graphics, images and tables included in the thesis.

Place, date

Asmik Nalmpatian