Chapters from modeling mortality in Hungary

Livia Varga

Doctoral School of Economics, Business and Informatics

Supervisors:

Dávid Burka, Ph.D.
assistant professor

László Kovács, Ph.D.
assistant professor

Corvinus University of Budapest
Doctoral School of Economics, Business and Informatics

Chapters from modeling mortality in Hungary

Dissertation

Livia Varga

Budapest, 2025

Table of contents

List of tables

List of figures

Introduction

Mortality projections are of particular interest to actuaries and demographers. While for actuaries the calculation of life expectancy is important for expected risk estimation, for demographers the use of mortality projection models can be important for population projections. The three main pillars of population projections are the prediction of mortality, fertility, and migration, which can be made either deterministically or stochastically. This thesis attempts to stochastically model and forecast mortality in Hungary, and aims to support the development of the population projection model of the Hungarian Central Statistical Office (HCSO). Population projections can support the policy making process, for example, in relation to the pension system, family policy, and employment policy. Thus, modeling mortality indirectly supports policy making. In this doctoral research, we used stochastic mortality models derived from the Lee–Carter (1992) model.

The study by Lee and Carter (1992) is a milestone in stochastic mortality modeling. The main advantage of this model is its simplicity: age-specific mortality rates can be predicted by considering only a few factors. It is sufficient to consider the number of deaths and the population by age and year to fit this mortality model. Estimating the model parameters does not require the quantification of past influences on mortality, such as medical and social factors. Over the past 30 years, many improved and modified versions of the Lee–Carter model have been developed. These models can be discussed within a framework, known as the generalized age–period–cohort (GAPC) model framework (see Hunt–Blake 2014, 2015, Villegas et al. 2018). From the Lee–Carter model family, we have selected both those used to forecast mortality in a single population, and those used to study the mortality of different populations together. Among the single- and multi-population models (see, e.g., Villegas et al. 2017), particular attention was paid to the application and numerical programming of the latter. We have also succeeded in fitting multi-population variants of notable mortality models that are not present in the international literature.

Improved versions of the Lee–Carter model include the multi-population models mentioned above (see, e.g., Carter–Lee 1992, Li–Lee 2005, Li 2013, Kleinow 2015, Li et al. 2016, Enchev et al. 2017, Wen et al. 2021). Modified versions also include those that use several time-varying indices rather than a single index to capture the trend in mortality (see, e.g., Booth et al. 2002, Renshaw–Haberman 2003). In some models, the age-

dependent coefficient of the mortality trend is parametric, i.e., a predefined function of age (see, e.g., Cairns et al. 2006, Plat 2009), while in other models it is nonparametric, i.e., estimated by the model (see, e.g., Lee–Carter 1992). There is also a modified version of the Lee–Carter model that considers the fact that the rate of decline in mortality by age group is not constant over time: it may accelerate for elders, and slow down for younger age groups in the long run (see Li–Gerland 2011, Li et al. 2013). This phenomenon is known as rotation. There are also model variants that include the cohort effect, which means that the change (improvement) in mortality is not independent of the year of birth (see, e.g., Renshaw–Haberman 2006, Yang et al. 2016). This research has considered all of these modifications, using the Lee–Carter model as a benchmark for comparing the different models.

Within the framework of this research, we focused on four topics, each of which can improve the modeling of mortality in Hungary. These four topics are: 1) **regional mortality** projection, 2) prediction of **mortality by cause of death**, 3) investigation of the time-dependence of age-varying parameters, i.e., **rotation** and their incorporation into the prediction of some mortality models, and 4) application of **tree-based machine learning methods** to mortality projections. For these research topics, we used different but related mortality models, always comparing their results and predictive accuracy, and selecting the best performing model. Different multi-population models are fitted to each topic, and the results are always compared to those of the Lee–Carter model.

The main objective of the doctoral research is to apply numerous mortality models to Hungarian data, and to support the methodological development of mortality forecasting in the HCSO. An important question was whether a model that fit the data better than the Lee–Carter model could be found by considering extensions such as multi-population models, rotation, and the cohort effect. Four research topics were defined for applying the Lee–Carter model extensions. This allowed the multi-population models to be tested using mortality data from different groups within the Hungarian population. Of the various mortality models fitted in relation to the four research themes, we always selected the one that proved to be the best fit. The following four research questions were defined in relation to the four themes:

Q1) In the HCSO, the population projection is based on the forecast of mortality, fertility, and migration at the county and regional level. In connection with this task, we have been engaged in fitting and forecasting various multi-population

mortality models to the Hungarian regional data. The main research question related to the **regional mortality projection** is *which of the selected multi-population models is best suited to forecast Hungarian territorial mortality rates by age and sex in the long term*.

Q2) As a second part of the research, we produced a **forecast by cause of death**. The multi-population mortality models can also be fitted to the data by main causes of death. The application of multi-population models to this type of data is an innovative approach. The multi-population model variants were developed by clustering the mortality indices of the selected single-population models that describe the improvement in mortality. In this context, an interesting research question is *whether creating the multi-population models based on clustering mortality indices are suitable for predicting cause-specific Hungarian mortality rates by age and sex*.

Q3) The **rotation of the age-dependent parameters** is also an important issue in long-term forecasting. Allowing for the time-varying age-dependent parameters can improve the prediction accuracy of multi-population models. For the selected mortality models, we have applied the rotation of age-dependent parameters in the forecasting process. The question is *how different the results of these variants would be from the original multi-population models*.

Q4) The use of machine learning techniques is an innovative approach to long-term mortality forecasting. We have used **tree-based machine learning methods** to improve the prediction of standard stochastic mortality models. The aim of creating combined models was to quantify the cohort effect using machine learning methods and improve the accuracy of standard model predictions. The related research question is *whether prediction accuracy can be improved by combining stochastic mortality models with tree-based methods*.

In the following, Chapter I provides an overview of the selected model variants. This literature review is followed by three main chapters: the research review, the description of mortality in Hungary and the presentation of our research results. The literature review mainly provides an overview of the mortality models in a generalized

model framework that were used in the research. Chapter II, the research review, describes the data and methods used. Chapter III provides a descriptive overview of Hungarian mortality characteristics in recent decades by sex, cause of death and region. Chapter IV presents the results and main findings of the four research topics. The conclusion of the thesis summarizes the research as a whole.

I. Literature review

In this chapter, we first introduce in detail the Lee–Carter mortality model, and then the generalized age–period–cohort framework. Within this framework, we present the single- and multi-population models that we fitted to the Hungarian data during the research. After that, we present the theoretical background of age effect rotation and our methodology. Finally, we discuss the tree-based machine learning methods we apply to extend the mortality models with the cohort effect.

I.1. The mortality model of Lee and Carter

The publication of Lee and Carter, written in 1992, is a notable study concerning stochastic mortality models. The authors analyzed the long-term time series of US mortality data, and prepared a forecast. In their parsimonious model, the logarithms of the central mortality rates depend on age- and time-varying parameters that should be estimated. The mortality index on the right-hand side of the model's initial equation is the long-term trend, which shows how mortality has changed over the selected period. Its coefficient depends only on age, so the trend in mortality change can be differentiated by age. The other age-dependent parameter in the equation is the baseline shape of mortality. The initial equation is:

$$ln\, m_{xt} = \alpha_x + \beta_x \kappa_t + \varepsilon_{xt}, \qquad \varepsilon_{xt} \sim N(0, \sigma^2) \tag{1}$$

where m_{xt} is the central mortality rate, which is the ratio of the number of deaths (D_{xt}) to the mid-year population, called central exposure (E^c_{xt}) at age x and in year t.[1] The error

[1] Exposure refers to exposure to risk, which is defined as the number of individuals in a specific age group and year who are at risk of dying.

term ε_{xt} is a white noise process with zero mean and constant variance (σ^2). In the equation, κ_t denotes the mortality index. The factor $\beta_x \kappa_t$ expresses how mortality by age has changed over time. Two constraints ensure that the solution is unique. These are the following: $\sum_x \beta_x = 1$ and $\sum_t \kappa_t = 0$. Under these constraints, the term α_x is equal to the average of $ln\, m_{xt}$ over time. The values of the mortality index κ_t always include both positive and negative values. The values of the coefficient β_x may be negative for some ages. This means that while mortality is increasing at some ages, it is decreasing at other ages.

To the right of the initial equation are unknown parameters that need to be estimated. Lee and Carter used singular value decomposition (SVD) for estimation. Using this method, the second step is to re-estimate κ_t. The following formula should be taken into account: $D_t = \sum\{exp\ (\alpha_x + \beta_x \kappa_t)\ E_{xt}\}$. After the re-estimation, the different sizes of the age groups result in different weights, instead of the earlier same weights. The assumption is that the trend observed in the past will continue in the future, and therefore a time series analysis and a prediction of the mortality index is needed to obtain the values of future mortality rates. Lee and Carter, using the Box–Jenkins model technique, found that κ_t is a random walk with a drift process. Lee–Miller (2001) proposed a modification of the projection in which α_x does not represent the average of the logarithms of mortality rates over the entire observation period by age. Instead of these values, the authors used the mortality rates of the last period in the projection to avoid the jump-off bias, as mortality rates in the last year of the base period can be much more favourable than the average for the observation period.

I.2. Some modifications of the Lee–Carter model

As mentioned above, several improved versions of the Lee–Carter model have been developed in recent years. However, a modified version has already been presented in the authors' paper. As the US mortality data were taken from 1900 onwards, they had to filter out the effect of the Spanish flu epidemic, which had a significant impact on mortality in 1918. For this reason, a dummy variable was used in the prediction of the mortality index. Criticisms of the Lee–Carter model are summarized by Lee (2000). One weakness of the model is that it assumes past patterns will continue into the future. However, this is not necessarily true, as structural breaks in the mortality index may occur (e.g., due to an

unexpected pandemic). If mortality improvement has been linear in the past, this does not mean that it will continue in the future. Another criticism raised by Lee (2000) is that the rate of decline in the mortality index may differ between age groups, which is not accounted for in the original Lee–Carter model. Lee (2000) mentioned that instead of the two-stage procedure, Wilmoth (1993) developed a one-stage process, the weighted SVD method. This uses the number of deaths for each age as a weight.

In the following, we first introduce the notations used for the mortality models in this thesis. The selected single- and multi-population models are then described in chronological order. These include models with multiple mortality indices and models with nonparametric age-dependent coefficients on the mortality index(es). We then introduce the phenomenon of rotation, which can be an important extension of mortality models for long-term projections. We describe separately the original method known from the literature and the possible solution for rotation that we have adopted.

I.2.1. Mortality models described within the framework of generalized age–period–cohort model

The publications of Hunt–Blake (2014, 2015) and Villegas et al. (2018) provide a summary of the original Lee–Carter model and some modified versions in a generalized framework. The study by Villegas et al. (2018) also introduces the R software package called *StMoMo*, which enables a number of notable mortality models to be fitted. However, we did not use it in our research. Instead, we applied our own numerical programming (see Chapter II). Following Villegas et al. (2018), this chapter introduces the generalized age–period–cohort framework.

First, the notations used in the research are introduced. D_{xtj} refers to the number of deaths at age x in year t in population j. E^c_{xtj} denotes the central exposure, i.e., the number of individuals at age x in the middle of year t in population j. E^0_{xtj} is the initial exposure, i.e., the number of individuals at age x at the beginning of year t in population j. The relationship between central and initial exposure can be described as follows: $E^0_{xtj} \approx E^c_{xtj} + \frac{1}{2} D_{xtj}$. Let q_{xtj} be the one-year probability of death, which can be estimated in the following way: $\hat{q}_{xtj} = D_{xtj}/E^0_{xtj}$. This is the initial mortality rate, which means the probability of death for an individual aged x in calendar year t. The central death rate m_{xtj} can be calculated as D_{xtj}/E^c_{xtj}. The conversion between initial and central

mortality rates can be described as follows: $q_{xtj} \approx 1 - e^{-m_{xtj}}$. Since it is assumed that the force of mortality[2] μ_{xtj} remains unchanged over age x and during year t, it can be concluded that μ_{xtj} and the central death rate m_{xtj} coincide. The empirical estimate of μ_{xtj} is: $\hat{\mu}_{xtj} = m_{xtj} = D_{xtj}/E^c_{xtj}$.

The structure of mortality models within the GAPC framework can be described by four components:

1. The *random component* refers to the distribution of the number of deaths D_{xtj}, i.e., Poisson or Binomial distribution.

$$D_{xtj} \sim Poisson(E^c_{xtj}\mu_{xtj}) \tag{2}$$

or

$$D_{xtj} \sim Binomial(E^0_{xtj},\ q_{xtj}) \tag{3}$$

where $\mathbb{E}\left(\frac{D_{xtj}}{E^c_{xtj}}\right) = \mu_{xtj}$ and $\mathbb{E}\left(\frac{D_{xtj}}{E^0_{xtj}}\right) = q_{xtj}$.

2. The *systematic component* describes the estimated equation. The factors affecting the intensity of mortality are age, period, and cohort.[3] Accordingly, the general form of the equation can be written as follows:

$$\eta_{xtj} = \alpha_{xj} + \sum_{i=1}^{n} \beta^{(i)}_{xj}\kappa^{(i)}_{tj} + \beta^{(0)}_{xj}\gamma_{(t-x)j} \tag{4}$$

where $n \geq 0$ is an integer, $\kappa^{(i)}_{tj}$ denotes the mortality index, which can be multiple in the equation, $i = 1, \ldots, n$. The coefficients $\beta^{(i)}_{xj}$ measure the effect of the mortality index(es) by age. The term $\gamma_{(t-x)j}$ means the cohort effect, and its age-varying coefficient is: $\beta^{(0)}_{xj}$. The $\beta^{(i)}_{xj}$ can be either parametric, i.e., predefined functions of age ($\beta^{(i)}_{xj} \equiv f^i(x)$), or nonparametric terms that do not need to be estimated in advance (e.g., in the Lee–Carter model). In the GAPC framework, we assume that $\kappa^{(i)}_{tj}$ and $\gamma_{(t-x)j}$ are stochastic processes.

[2] In other words, the instantaneous death rate or hazard rate. See Thatcher et al. (1998).
[3] The cohort can be calculated as follows: $c = t - x$.

3. The *link function* g is a log function for the Poisson distribution and a logit function for the Binomial distribution of deaths.

$$g\left(\mathbb{E}\left(\frac{D_{xtj}}{E_{xtj}}\right)\right) = \eta_{xtj} \tag{5}$$

4. A *set of parameter constraints* is required to ensure a unique solution. The parameters are:

$$\theta \coloneqq (\alpha_{xj}, \beta_{xj}^{(1)}, \dots, \beta_{xj}^{(n)}, \kappa_{tj}^{(1)}, \dots, \kappa_{tj}^{(n)}, \beta_{xj}^{(0)}, \gamma_{(t-x)j}). \tag{6}$$

In the following, the selected modified versions of the Lee–Carter model are presented using the above notations.

I.2.2. Single-population mortality models

The Lee–Carter model with Poisson and Binomial distributions

In the Lee–Carter model, the error term is homoscedastic and normally distributed. However, this is not realistic because the logarithms of the mortality rates are more variable in the older age groups than in the younger groups. Brouhns et al. (2002) assumed a Poisson distribution of deaths, and mentioned that, according to Brillinger (1986), this is the most appropriate assumption for the number of deaths. Renshaw–Haberman (2003) also considered a Poisson distribution. The model proposed by Brouhns et al. (2002) can be fitted using a log function and central mortality rates. Similar to the original Lee–Carter model, a static age function (α_{xj}) and a nonparametric age–period factor ($\beta_{xj}\kappa_{tj}$) appear on the right side of the equation. Thus, the model has a mortality index. Brouhns et al. (2002) did not include the cohort effect in their model. The initial equation is therefore:

$$\eta_{xtj} = \alpha_{xj} + \beta_{xj}^{(1)}\kappa_{tj}^{(1)}. \tag{7}$$

Brouhns et al. (2002) used the same constraints as Lee and Carter, but a scaling constraint was also applied to the age effect. Using our notation, the constraints are as follows: $\sum_x \beta_{xj}^{(1)} = 1$, $\beta_{1j}^{(1)} = 1$, and $\sum_t \kappa_{tj}^{(1)} = 0$ for $\forall\, j$. In order to forecast the mortality rates after fitting the parameters, the mortality index should be forecasted. According to Lee–Carter (1992), the mortality index follows a random walk with a drift process, i.e., ARIMA(0,1,0):[4]

$$\kappa_{tj}^{(1)} = \delta + \kappa_{(t-1)j}^{(1)} + \epsilon_{tj}, \quad \epsilon_{tj} \sim N(0, \sigma_\kappa^2) \tag{8}$$

where δ is the drift parameter, and ϵ_{tj} is white noise. Brouhns et al. (2002) used maximum likelihood estimation with Newton–Raphson iteration. Instead of the ARIMA(0,1,0) model, Brouhns et al. (2002) fitted ARIMA(0,1,1) to forecast the time-varying index. The R package *StMoMo*, developed by Villegas et al. (2018), provides the ability to fit the Lee–Carter model with a log or logit link function and under the assumption of a Poisson or Binomial distribution.

The Lee–Carter model with two time-varying mortality indices

Booth et al. (2002) and Renshaw–Haberman (2003) incorporated two (or more) time-varying indices with age-dependent coefficients into the mortality model. The original Lee–Carter model uses only the first terms of the SVD. However, systematic variation may remain in the residuals if the model has only one mortality index. The second (or more) bilinear component aims to capture possible temporal effects that cannot be captured by one index. The initial equation of the mortality model with two time-varying factors can be written as follows:

$$\eta_{xtj} = \alpha_{xj} + \beta_{xj}^{(1)}\kappa_{tj}^{(1)} + \beta_{xj}^{(2)}\kappa_{tj}^{(2)}. \tag{9}$$

[4] ARIMA is the short name of the autoregressive integrated moving average models.

The following constraints are necessary for the unique solution: $\sum_x \beta_{xj}^{(1)} = 1$, $\sum_x \beta_{xj}^{(2)} = 1$, $\sum_t \kappa_{tj}^{(1)} = 0$, and $\sum_t \kappa_{tj}^{(2)} = 0$ for $\forall\, j$.[5] We can use multivariate vector time series methods or ARIMA models separately to forecast mortality indices. This model was also used in the research under the assumptions of both Poisson and Binomial distributions.

The Cairns–Blake–Dowd (CBD) model

Cairns et al. (2006) also introduced a two-factor stochastic model. The CBD model includes two mortality indices. The first index affects the dynamics of mortality rates in the same way for all ages, while the second factor has a much stronger effect on older age groups than on younger cohorts. The coefficients $\beta_{xj}^{(i)}$ are: $\beta_{xj}^{(1)} = 1$ and $\beta_{xj}^{(2)} = x - \bar{x}$ for $\forall\, j$. There is no cohort effect and no α_{xj} term in the CBD model. Villegas et al. (2018) assumed a Binomial distribution of deaths, and applied a logit link function[6] with respect to this model.

$$\eta_{xtj} = \kappa_{tj}^{(1)} + (x - \bar{x})\kappa_{tj}^{(2)} \tag{10}$$

where $\bar{x}$ is the mean age. We can use a bivariate random walk with drift to predict $\kappa_{tj}^{(1)}$ and $\kappa_{tj}^{(2)}$. Another way of forecasting mortality indices is to fit ARIMA models separately. No identification problem arises in the CBD model making constraints unnecessary. This model is particularly appropriate for older age groups. Cairns et al. (2009) extended this model to include a cohort effect. Cairns et al. (2009, 2011) and Dowd et al. (2020) considered even further extensions of the CBD model. Overall, an entire family of CBD models have been developed.

The reduced versions of the Plat model

Plat (2009) attempted to create a model that retains the advantages of existing mortality models, while eliminating their disadvantages. Plat's model combines the CBD model

[5] Renshaw–Haberman (2003) applied the four constraints. The study by Hunt–Blake (2020) presents a total of six identifiability constraints in order to obtain a fully identified model.

[6] $logit(q_{xtj}) = ln\left(\frac{q_{xtj}}{1-q_{xtj}}\right) = \eta_{xtj}$

with some features of the Lee–Carter model to provide an equation that is suitable for all ages, and includes the cohort effect. Plat (2009) assumed a Poisson distribution and a log link function. There are three mortality indices, a cohort effect and an α_{xj} term in the initial equation:

$$\eta_{xtj} = \alpha_{xj} + \kappa_{tj}^{(1)} + (\bar{x} - x)\kappa_{tj}^{(2)} + (\bar{x} - x)^{+}\kappa_{tj}^{(3)} + \gamma_{(t-x)j}. \quad (11)$$

In this model, $\beta_{xj}^{(0)} = 1$, $\beta_{xj}^{(1)} = 1$, $\beta_{xj}^{(2)} = \bar{x} - x$, and $\beta_{xj}^{(3)} = (\bar{x} - x)^{+} = \max(0, \bar{x} - x)$ for $\forall\, j$. There are four stochastic factors $(\kappa_{tj}^{(1)}, \kappa_{tj}^{(2)}, \kappa_{tj}^{(3)}, \gamma_{(t-x)j})$ in the full Plat model. For the mortality indices, the following constraints are necessary: $\sum_t \kappa_{tj}^{(1)} = 0$, $\sum_t \kappa_{tj}^{(2)} = 0$, and $\sum_t \kappa_{tj}^{(3)} = 0$ for $\forall\, j$. Additional constraints are: $\sum_{c=t_1-x_k}^{t_n-x_1} \gamma_{cj} = 0$, $\sum_{c=t_1-x_k}^{t_n-x_1} c\, \gamma_{cj} = 0$, and $\sum_{c=t_1-x_k}^{t_n-x_1} c^2\, \gamma_{cj} = 0$ where $c = t_1 - x_k$ is the earliest, and $t_n - x_1$ is the latest year of birth. The first three constraints guarantee that the mortality indices will fluctuate around zero. The other three equations imply that the cohort effect will also fluctuate around zero without a linear or quadratic trend.

Excluding the cohort effect[7] the initial equation mentioned above is:

$$\eta_{xtj} = \alpha_{xj} + \kappa_{tj}^{(1)} + (\bar{x} - x)\kappa_{tj}^{(2)} + (\bar{x} - x)^{+}\kappa_{tj}^{(3)}. \quad (12)$$

Our reduced model includes three stochastic factors ($\kappa_{tj}^{(1)}$, $\kappa_{tj}^{(2)}$, $\kappa_{tj}^{(3)}$). $\kappa_{tj}^{(1)}$ represents the change in mortality at all ages, and $\kappa_{tj}^{(2)}$ allows the change in mortality to be varied by age. $\kappa_{tj}^{(3)}$ captures that the dynamics of mortality rates may be different in younger age groups (e.g., due to drug and alcohol use). The three constraints on the mortality indices are the same as above.

According to Plat (2009), the ARIMA(0,1,0) model is the best choice for the $\kappa_{tj}^{(1)}$ factor, and ARIMA(1,0,0) without intercept is an appropriate model for $\kappa_{tj}^{(2)}$, $\kappa_{tj}^{(3)}$, and $\gamma_{(t-x)j}$ with respect to the original model. If we are only interested in the older age group,

[7] As the fitted values of the cohort effect tend to be quite volatile, adding uncertainty to the forecast, the cohort effect is excluded from the models in this research.

the factor $\kappa_{tj}^{(3)}$ can be omitted from the model. The initial equation for the second version of our reduced Plat model is:

$$\eta_{xtj} = \alpha_{xj} + \kappa_{tj}^{(1)} + (\bar{x} - x)\kappa_{tj}^{(2)}. \quad (13)$$

The constraints on $\kappa_{tj}^{(1)}$ and $\kappa_{tj}^{(2)}$ are the same as above.

Model summary

Table 1 summarizes the above single-population mortality models used in the doctoral research. For single-population models, all estimated parameters are subpopulation-specific. The Poisson and Binomial Lee–Carter models with two factors are the most complex of the selected single-population models in the sense that they have the largest number of parameters to be estimated. As with this model, the Lee–Carter model was fitted under both distributional assumptions. The reduced Plat model with three factors includes three mortality indices, i.e., most of the stochastic parameters should be predicted in this case.

The CBD model is the most parsimonious in this research because only two mortality indices need to be fitted, and the coefficients of the indices are parametric, fixed in advance. The Lee–Carter model and the Plat model with two factors also have a small number of parameters to estimate. In both models there are three. Initial exposure and logit link function were used for models assuming a Binomial distribution for the number of deaths. Central exposure and log link function were applied in relation to models assuming a Poisson distribution of deaths. For the CBD model, no constraints are required, and for the reduced Plat models, constraints are required only for time-dependent parameters.

Table 1. Summary of the selected single-population stochastic mortality models

The names of the models	The initial equations	Constraints	Literature
Poisson Lee–Carter*	$ln\, m_{xtj} = \alpha_{xj} + \beta_{xj}^{(1)}\kappa_{tj}^{(1)}$	$\sum_x \beta_{xj}^{(i)} = 1$ and $\sum_t \kappa_{tj}^{(i)} = 0$ for $\forall\, i, j$ $i = 1, 2$	Lee–Carter (1992), Brouhns et al. (2002), Renshaw–Haberman (2003)
Binomial Lee–Carter**	$ln\left(\frac{q_{xtj}}{1-q_{xtj}}\right) = \alpha_{xj} + \beta_{xj}^{(1)}\kappa_{tj}^{(1)}$		Lee–Carter (1992), Villegas et al. (2018)
Poisson Lee–Carter with two factors*	$ln\, m_{xtj} = \alpha_{xj} + \beta_{xj}^{(1)}\kappa_{tj}^{(1)} + \beta_{xj}^{(2)}\kappa_{tj}^{(2)}$		Booth et al. (2002), Renshaw–Haberman (2003)
Binomial Lee–Carter with two factors**	$ln\left(\frac{q_{xtj}}{1-q_{xtj}}\right) = \alpha_{xj} + \beta_{xj}^{(1)}\kappa_{tj}^{(1)} + \beta_{xj}^{(2)}\kappa_{tj}^{(2)}$		Own modification based on the previous model.
CBD**	$ln\left(\frac{q_{xtj}}{1-q_{xtj}}\right) = \kappa_{tj}^{(1)} + (x-\bar{x})\kappa_{tj}^{(2)}$	–	Cairns et al. (2006), Villegas et al. (2018)
reduced Plat with three factors*	$ln\, m_{xtj} = \alpha_{xj} + \kappa_{tj}^{(1)} + (\bar{x}-x)\kappa_{tj}^{(2)} + (\bar{x}-x)^{+}\kappa_{tj}^{(3)}$	$\sum_t \kappa_{tj}^{(i)} = 0$ for $\forall\, i, j$ $i = 1, 2, 3$	Own modification based on the basis of Plat (2009).
reduced Plat with two factors*	$ln\, m_{xtj} = \alpha_{xj} + \kappa_{tj}^{(1)} + (\bar{x}-x)\kappa_{tj}^{(2)}$		Own modification based on the basis of Plat (2009).

Notes: The subscript j refers to the subpopulation, $\bar{x}$ is the mean age, and $(\bar{x}-x)^{+} = \max(0, \bar{x}-x)$. Central exposure was used for models marked with * and initial exposure for models marked with **.
Source: own editing

I.2.3. Multi-population mortality models

This chapter deals with multi-population models belonging to the Lee–Carter model family. If we would like to model the mortality of populations with similar socioeconomic backgrounds, the choice of multi-population models may be the most appropriate. The

study by Villegas et al. (2017) introduces several multi-population models by other authors (see, e.g., Carter–Lee 1992, Li–Lee 2005, Kleinow 2015, Yang et al. 2016). Subgroups of a population may be, for example, different countries of a region or territorial units of a country, or groups disaggregated by sex.

In the long run, projecting mortality by subpopulation separately may increase the divergence between groups in terms of age-specific mortality rates and life expectancy. For example, infant mortality rates for girls tend to be lower than for boys in developed countries, and it is implausible to assume the opposite in the future. Avoiding the crossover effect has led to the development of multi-population mortality models, which allow subpopulations to be analyzed together, assuming a close relationship between them. If a multi-population model avoids this divergence problem, it is a coherent model (see Li–Lee 2005). Not all multi-population models produce coherent forecasts. These models have both group-specific and common parameters in their initial equation. In the following, we present the multi-population models that we fitted to Hungarian data as part of the doctoral research. We use the notation described above (see Chapter I.2.1).

The Carter–Lee model

Carter and Lee (1992) analyzed sex-disaggregated data using a model with an initial equation that differs from the Lee–Carter model in only one parameter: the mortality index is the same for both subgroups, all other parameters are group-specific. The initial equation of the Carter–Lee model can be written as follows:

$$\eta_{xtj} = \alpha_{xj} + \beta_{xj}^{(1)} \kappa_t^{(1)} \tag{14}$$

where j refers to the subpopulation. The necessary constraints are: $\sum_x \beta_{xj}^{(1)} = 1$ for $\forall\, j$ and $\sum_t \kappa_t^{(1)} = 0$. However, this model does not necessarily produce a coherent forecast.

The common factor (CF) model

Li and Lee (2005) mentioned a version of the Lee–Carter model where the divergence in mortality between subpopulations is no longer an issue. In this model, the mortality index

and its coefficient are common to the subpopulations. The initial equation of this multi-population model is:

$$\eta_{xtj} = \alpha_{xj} + \beta_x^{(1)}\kappa_t^{(1)} \tag{15}$$

where the $\beta_x^{(1)}\kappa_t^{(1)}$ term is referred to as a common factor. The constraints are as follows: $\sum_x \beta_x^{(1)} = 1$ and $\sum_t \kappa_t^{(1)} = 0$. The forerunner of this model was the research of Lee–Nault (1993), who analyzed spatial data from Canada. In order to avoid the problem of divergence, they had already applied the common factor model. The model is particularly suitable when $\beta_x^{(1)}$ does not differ significantly by subpopulation. The term α_{xj} is the only group-specific factor in this model, but it does not cause a divergence problem. The authors originally used the SVD method to fit the parameters.

The augmented common factor (ACF) model

An additional time-varying factor may be included in the CF model to account for short- and medium-term deviations from the long-term mortality trend. The additional index and its coefficient may differ between groups. This variant of the model is also presented by Li–Lee (2005). The initial equation is:

$$\eta_{xtj} = \alpha_{xj} + \beta_x^{(1)}\kappa_t^{(1)} + \beta_{xj}^{(2)}\kappa_{tj}^{(2)}. \tag{16}$$

In this model, the common factor represents the long-term trend in mortality that is common to the subpopulations, and the second factor represents the shorter-term deviations that are specific to each population j. In a first step, Li and Lee fit the original Lee–Carter model to the aggregated data, and in a next step, the additional factor was estimated on the residual matrix[8] using the SVD method. The constraints on the parameters are: $\sum_x \beta_x^{(1)} = 1$, $\sum_t \kappa_t^{(1)} = 0$, $\sum_x \beta_{xj}^{(2)} = 1$, and $\sum_t \kappa_{tj}^{(2)} = 0$ for $\forall\, j$.

When predicting the factors $\kappa_t^{(1)}$ and $\kappa_{tj}^{(2)}$, we can assume independence between them, as mentioned by Li–Lee (2005). The independence of the group-specific second

[8] $ln\, m_{xtj} - \alpha_{xj} - \beta_x^{(1)}\kappa_t^{(1)}$

factors can also be assumed, as these deviations from the long-term trend are random for different subpopulations. According to Li–Lee (2005), a random walk with a drift can be used to forecast mortality rates for both common and group-specific mortality indices. The authors also mentioned the AR(1) model for the second index.

The augmented common factor (ACF) model with more factors

Li (2013) modified the ACF model in two aspects. On the one hand, a Poisson distribution was assumed for the number of deaths, and the author estimated the model using the maximum likelihood method. On the other hand, there was more than one additional group-specific factor in the model of Li (2013). The initial equation of the ACF model with more factors is:

$$\eta_{xtj} = \alpha_{xj} + \beta_x^{(1)}\kappa_t^{(1)} + \sum_{i=2}^{n}\beta_{xj}^{(i)}\kappa_{tj}^{(i)}. \quad (17)$$

However, fitting a model with too many parameters should be avoided, especially if the third- or higher-order indices appear as irregular components. In this case, the number of additional factors should be reduced. Li (2013) distinguished at most six additional indices, and suggested AR(p) models for forecasting the additional time-varying mortality indices. In addition, Li (2013) used a random walk with a drift model for the common mortality index. Regarding this model, the following constraints are applied: $\sum_x \beta_x^{(1)} = 1$, $\sum_t \kappa_t^{(1)} = 0$, $\sum_x \beta_{xj}^{(i)} = 1$, and $\sum_t \kappa_{tj}^{(i)} = 0$ for $\forall\, i, j$.

The common age effect (CAE) model

Kleinow (2015) proposed the use of the common age-varying coefficient, in addition to the group-specific time-dependent parameter, to construct a multi-population model. According to Kleinow's modification, the initial equation of the Lee–Carter model changes as follows:

$$\eta_{xtj} = \alpha_{xj} + \beta_x^{(1)}\kappa_{tj}^{(1)} \quad (18)$$

where the mortality index is group-specific. The necessary constraints are: $\sum_x \beta_x^{(1)} = 1$ and $\sum_t \kappa_{tj}^{(1)} = 0$ for $\forall\, j$. Kleinow (2015) estimated a common age effect using central mortality rates and the common principal component analysis method. Kleinow argued for estimation methods that use mortality rates, which is the opposite of the maximum likelihood method. The latter calculates directly with the number of observed deaths and the population. Thus, the common age effect ($\beta_x^{(1)}$) depends significantly on the mortality of larger populations.

The augmented common factor model with common age effect (ACF–CAE)

Following the idea of Kleinow (2015), Li et al. (2016) mentioned the possibility of using common age-varying coefficients with respect to an augmented common factor model with more factors. This variant is called the augmented common factor model with common age effect (ACF–CAE). The initial equation can be written as follows:

$$\eta_{xtj} = \alpha_{xj} + \beta_x^{(1)}\kappa_t^{(1)} + \sum_{i=2}^{n} \beta_x^{(i)}\kappa_{tj}^{(i)}. \tag{19}$$

The related constraints are: $\sum_x \beta_x^{(1)} = 1$, $\sum_t \kappa_t^{(1)} = 0$, $\sum_x \beta_x^{(i)} = 1$, and $\sum_t \kappa_{tj}^{(i)} = 0$ for $\forall\, i, j$. If the group-specific age effects $\beta_{xj}^{(i)}$ are very similar for the subpopulations, it is particularly important to include common age-varying coefficients in the initial equation of the mortality model.

Model summary

Table 2 provides an overview of the selected multi-population mortality models. The ACF model with three factors is the most complex multi-population model in this research, with seven parameters to fit based on the current time series. A multi-population version of the Lee–Carter model is the Carter–Lee model, which, like the single-population version, was fitted under both distribution assumptions. The models with the most stochastic parameters are ACF with three factors and the multi-population version of the reduced Plat with three factors. The multi-population variant of the CBD model is the most parsimonious, as the other models require more than two parameters to be estimated.

Initial exposure and the logit link function were used for the Binomial Carter–Lee model, the multi-population version of the Binomial Lee–Carter model with two factors, and the CBD model with multi-population extension. No constraints were necessary in the latter case. For all other models, we used a log link function and assumed a Poisson distribution. Of the selected multi-population models, the Poisson/Binomial Carter–Lee and CF models are those where the mortality index is not group-specific but common to all subgroups. The CAE model includes a group-specific mortality index. All other models include both a group-specific and a common mortality index.

One advantage of multi-population models is that they enable the mortality of closely related subpopulations to be analyzed jointly. Some of these models allow for coherent forecasting of different subgroups. Furthermore, the number of parameters that need to be forecasted can be reduced since multi-population models incorporate common factors, which do not need to be predicted for each group. Another advantage of multi-population models with multiple mortality indices is that they allow us to examine group-specific, shorter-term variation from a long-term trend for the whole population by taking into account common and group-specific time-dependent factors. The drawback is that coherent forecasting is not guaranteed by all multi-population models. Another drawback is that, when examining subpopulations with similar mortality trends but different intensities of mortality, the group with the higher mortality rate exerts a stronger influence on the fitted parameters in multi-population models.

Table 2. Summary of the multi-population stochastic mortality models

The names of the models	The initial equations	Constraints	Literature
Poisson Carter–Lee*	$ln\, m_{xtj} = \alpha_{xj} + \beta_{xj}^{(1)}\kappa_t^{(1)}$	$\sum_x \beta_{x(j)}^{(i)} = 1$ and $\sum_t \kappa_{t(j)}^{(i)} = 0$ for $\forall\, i, j$ $i = 1, 2$	Carter–Lee (1992), Lee–Nault (1993)
Binomial Carter–Lee**	$ln\left(\frac{q_{xtj}}{1-q_{xtj}}\right) = \alpha_{xj} + \beta_{xj}^{(1)}\kappa_t^{(1)}$		Own modification based on the previous model.
CF*	$ln\, m_{xtj} = \alpha_{xj} + \beta_{x}^{(1)}\kappa_t^{(1)}$		Li–Lee (2005)
ACF*	$ln\, m_{xtj} = \alpha_{xj} + \beta_{x}^{(1)}\kappa_t^{(1)} + \beta_{xj}^{(2)}\kappa_{tj}^{(2)}$		Li–Lee (2005), Li (2013)
Poisson Lee–Carter with two factors (multi-population extension)*	$ln\, m_{xtj} = \alpha_{xj} + \beta_{xj}^{(1)}\kappa_{tj}^{(1)} + \beta_{xj}^{(2)}\kappa_t^{(2)}$		Own modifications. These multi-population models are based on the single-population models. See Table 1 for related literature.
Binomial Lee–Carter with two factors (multi-population extension)**	$ln\left(\frac{q_{xtj}}{1-q_{xtj}}\right) = \alpha_{xj} + \beta_{xj}^{(1)}\kappa_{tj}^{(1)} + \beta_{xj}^{(2)}\kappa_t^{(2)}$		
CBD (multi-population extension)**	$ln\left(\frac{q_{xtj}}{1-q_{xtj}}\right) = \kappa_t^{(1)} + (x-\bar{x})\kappa_{tj}^{(2)}$	–	
reduced Plat with three factors (multi-population extension)*	$ln\, m_{xtj} = \alpha_{xj} + \kappa_t^{(1)} + (\bar{x}-x)\kappa_{tj}^{(2)} + (\bar{x}-x)^{+}\kappa_{tj}^{(3)}$	$\sum_t \kappa_{t(j)}^{(i)} = 0$ for $\forall\, i, j$ $i = 1, 2, 3$	
reduced Plat with two factors (multi-population extension)*	$ln\, m_{xtj} = \alpha_{xj} + \kappa_t^{(1)} + (\bar{x}-x)\kappa_{tj}^{(2)}$		
ACF with three factors*	$ln\, m_{xtj} = \alpha_{xj} + \beta_{x}^{(1)}\kappa_t^{(1)} + \beta_{xj}^{(2)}\kappa_{tj}^{(2)} + \beta_{xj}^{(3)}\kappa_{tj}^{(3)}$	$\sum_x \beta_{x(j)}^{(i)} = 1$ and $\sum_t \kappa_{t(j)}^{(i)} = 0$ for $\forall\, i, j$ $i = 1, 2, 3$	Li (2013)
CAE*	$ln\, m_{xtj} = \alpha_{xj} + \beta_{x}^{(1)}\kappa_{tj}^{(1)}$		Kleinow (2015)
ACF–CAE*	$ln\, m_{xtj} = \alpha_{xj} + \beta_{x}^{(1)}\kappa_t^{(1)} + \beta_{x}^{(2)}\kappa_{tj}^{(2)}$		Li et al. (2016)

Notes: The subscript j refers to the subpopulation, $\bar{x}$ is the mean age, and $(\bar{x}-x)^{+} = \max(0, \bar{x}-x)$. Central exposure was used for models marked with * and initial exposure for models marked with **.

Source: own editing

I.2.4. The rotation of the age pattern

The Lee–Carter method extended with a rotation model

In developed countries, mortality improvement slows down in younger age groups, and accelerates in older cohorts. Li et al. (2013) called this phenomenon rotation. For infants and younger age groups, a rapid rate of improvement cannot be expected in the long term, as this would lead to unreasonably low mortality rates. Based on the past experience, the age shape of the mortality is U-shaped, with relatively high mortality among infants, very low mortality among young people and increasing mortality among the elderly. This is not expected to change in the long term. In the case of the elderly, the acceleration of mortality improvement may be justified by future medical breakthroughs. The older age group can make a greater contribution to increasing life expectancy in the future. Li et al. (2013) expected that rotation will also occur in developing countries as they achieve higher life expectancy. Li–Gerland (2011) developed the Lee–Carter method with robust rotation. This differs from the model of Li et al. (2013). In the following, we present the latter model in more detail.

The model of Li et al. (2013) assumed that the β_x coefficient in the original Lee–Carter model is not constant in the long run. Nevertheless, they mentioned that the Lee–Carter model may be appropriate in the medium term with the unrotated β_x, as it may be stable over a period of time. However, if the rotation is not considered, the results of the long-term forecast become questionable. The authors have modified the shape of the β_x curve. They highlighted the mortality rates for infants and for 15–19-year-olds. Mortality is among the lowest for the latter age group and higher for infants. The ratio of the mortality rates of the two groups is greater than one. The results of the rotation were evaluated in terms of this ratio. For the selected 20 countries with the lowest mortality between 2005 and 2010, a decreasing trend in the value of the ratio m_0/m_{15-19} was observed. Taking this trend into account, when life expectancy reaches a certain level, the β_x coefficient in the Li et al. (2013) model will gradually decrease, and the aforementioned ratio will follow this downward trend. As the sum of β_x is equal to one, a decrease in the values of this parameter in younger age groups is automatically associated with an increase in older age groups.

Li et al. (2013) introduced the following notation: $e_0(t)$ is the life expectancy in year t for men and women together. This can be estimated using the Lee–Carter model or

other similar models can be applied. The $b_0(x)$ refers to the age-specific pattern of mortality improvement at age x. This is β_x from the original Lee–Carter model from which we start the rotation. The $b_u(x)$ is the assumed ultimate value. The e_0^u sign indicates the final life expectancy at the end of the rotation. $B(x,t)$ is the age pattern of mortality in the extended Lee–Carter model, which has a linear-weight and a smooth-weight function:

$$w(t) = \frac{[e_0(t)-80]}{[e_0^u-80]}, \tag{20}$$

$$w_s(t) = \left\{0.5\left[1 + sin\left[\frac{\pi}{2}(2w(t)-1)\right]\right]\right\}^p. \tag{21}$$

Then $B(x,t)$ is equal to the following values:

$$B(x,t) = \begin{cases} b_0(x), & e_0(t) < 80, \\ \big(1-w_s(t)\big)b_0(x) + w_s(t)b_u(x), & 80 \le e_0(t) < e_0^u, \\ b_u(x), & e_0^u \le e_0(t). \end{cases} \tag{22}$$

The value of the linear-weight function is zero when $e_0(t)$ equals 0, and one when $e_0(t) = e_0^u$. The smooth-weight function results in a continuous change in mortality rates. The p value is between zero and one. This parameter makes the rotation faster at first, and then slower over time. If $p = 1$, there is no smoothing of the mortality decline pattern by age. The default value is 0.5. The age of 80 was approximately the median life expectancy in 2008 based on data from the 20 countries analyzed. We can see that if $e_0(t)$ is less than 80, then $B(x,t) = b_0(x)$.

The rotation is realized when $80 \le e_0(t) < e_0^u$, and when the $e_0(t)$ reaches the e_0^u. The rotation ends when $B(x,t) = b_u(x)$. A smaller value of e_0^u results in a faster rotation and a higher ratio m_0/m_{15-19} at the end of the forecast horizon. However, if the value of e_0^u is chosen high, this will result in a smoothed rotation and a smaller ratio m_0/m_{15-19} at the end of the projection, which may be even more questionable. Finally, Li et al. (2013) suggested an age of 102 for the value of e_0^u. The ratio m_0/m_{15-19} is decreasing, and this is expected to continue in the future.

The extended Lee–Carter model of Li et al. (2013) reduces to the original Lee–Carter model when the projected $e_0(t)$ is less than 80. When $e_0(t)$ is greater than 80, the

extended model gradually diverges from the original Lee–Carter model, as the improvement in mortality rates slows down for the younger age groups, and accelerates for the older ones. If life expectancy in a country is less than 80 years, the rotation takes place in the later years of the projection. If actual life expectancy is greater than 80 years, the rotation takes place throughout the projection period.

If we ignore the rotation of the age pattern, an important consequence may be that we underestimate the old-age population. The study by Vékás (2018) was the first to present the phenomenon of rotation on Hungarian data, following the methodology of Li et al. (2013). Vékás (2018) showed the importance of rotation using data for 22 age groups between 1950 and 2015: the rotation of mortality improvement was weakly detectable for Hungarian men, and significant for Hungarian women. Vékás (2020) examined the rotation for the Member States of the European Union, and showed that the rotation of the age effect is important in 11 of the 28 Member States for both sexes.

Changes over time in age-dependent parameters of mortality models

The study of the time-dependence of the parameter $\beta_{x(j)}^{(1)}$ and its rotation is widely used in the literature for two models. Typically, the Lee–Carter model and a multi-population version of it, the common factor model developed by Li–Lee (2005), use rotation in forecasting. However, there are models with multiple mortality indices and multiple age-varying parameters for which it may be worthwhile to consider rotation. Furthermore, not only the age effect(s) but also the baseline shape of mortality may vary over time. The term α_{xj} is also a static parameter in the forecast, but its value may decrease as mortality improves. Lee–Miller (2001) proposed a modification of the term α_{xj} at the forecast horizon. This means that the log mortality values for the most recent year observed are the baseline shape of mortality during the projection. The change in term α_{xj} over time may also be worth considering in mortality prediction models. In the following, we propose a simple method that can be used to investigate the dependence of all age-varying parameters on time, and to incorporate them into various mortality models.

In this thesis, we examined the change in age-varying parameters over time for three mortality models, the Lee–Carter, ACF and ACF with three factors models, and attempted to incorporate this into the prediction. We also use the term rotation for the

method used. The change in the values of the age-varying parameters over time can be estimated by the method of least squares as follows (see Varga 2024):

$$\alpha_{xtj} = a_{xj} + b_{xj}t + \epsilon_{\alpha,xtj}, \qquad \epsilon_{\alpha,xtj} \sim N(0, \sigma_\alpha^2) \tag{23}$$

and

$$\beta_{xt(j)}^{(i)} = c_{x(j)}^{(i)} + d_{x(j)}^{(i)}t + \epsilon_{\beta,xt(j)}^{(i)}, \qquad \epsilon_{\beta,xt(j)}^{(i)} \sim N(0, \sigma_\beta^2) \tag{24}$$

where the error terms $\epsilon_{\alpha,xtj}$ and $\epsilon_{\beta,xt(j)}^{(i)}$ are normally distributed. In order to see the change in age-dependent parameters over time for the current data, the mortality models have to be fitted to several base periods of different lengths. Thus, in Equations 23–24, t is equal to the number of base periods of different lengths. The parameters a_{xj}, b_{xj}, $c_{x(j)}^{(i)}$, and $d_{x(j)}^{(i)}$ estimated by the Equations 23–24 are used to predict the future values of α_{xtj} and $\beta_{xt(j)}^{(i)}$.

Due to the rotation, the constraint on $\beta_{xt(j)}^{(i)}$ is not violated as its sum over the years remains one in each year of the forecast. The reason is that, for each age year, the value of the parameter changes by a constant value from year to year over the forecast horizon. When we sum the value of change by age, we get zero: while the change is an increase at some ages, it is a decrease at others. Chapter IV.3 introduces the application of the rotation method described in Equations 23–24.

I.3. Tree-based machine learning methods

Machine learning methods are becoming increasingly popular in many research areas, and this is no exception in the field of mortality prediction. The application of machine learning methods to mortality projections is the fourth topic for the doctoral research due to its innovative nature. This chapter presents the selected tree-based methods, chosen according to Bjerre (2022). Tree-based models were applied to improve the forecasts of standard stochastic mortality prediction models. The decision tree (see Breiman et al. 1984) is applicable to both classification and regression problems. The latter is relevant for our purposes, and will be discussed here. When using machine learning methods, the observed period is usually divided into two parts: training and test data sets. The model

is trained on a training data set, a prediction is made for the test period, and the model is evaluated based on the accuracy of the prediction.

The essence of the decision tree is that the predictor space is divided into regions. The division into J distinct, nonoverlapping regions (R_J) is based on p explanatory variables (X_p). For the decision tree, the split is always based on the cutpoint of an explanatory variable. The same prediction is made for the observations i that fall in the R_J region. The predicted value for the R_J region is the average of the response values for the observations in that region. The purpose of the method is to find regions that minimize the residual sum of squares (RSS) (see James et al. 2013):

$$\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \tag{25}$$

where $\hat{y}_{R_j}$ denotes the mean response value for the training observations in the R_J region, and y_i are the current values for the observations i. First, each observation is assigned to a single region, and the cut is made from the top of the tree down. With each cut, the method separates two additional regions. This process is called recursive binary splitting. The cut is always made based on the best predictor and its best cutpoint to achieve the maximum reduction in RSS. Only one of the two regions produced by a cut is split further. As a result, the decision tree ends in leaves (or terminal nodes).

Overfitting the model means that a good fit is found for the training period, but the test set performance is poor. The validation approach is the right way to counteract overfitting. The disadvantage of the decision tree is that it is not robust to the results obtained. A small change in the data can have a large effect on the estimated tree. Considering the results obtained, the estimates of the decision tree suffer from high variance. However, using many decision trees (methods such as bagging, random forests or gradient boosting) can reduce variance, and increase prediction accuracy.

Random forests, which consist of multiple decision trees, can be attributed to Breiman (2001), who developed the decision tree in two ways. A feature of the method is the random selection of training data: bootstrap samples are selected for each decision tree. A number B of samples are selected from the training data set, and the method is trained on these samples. In addition, the variables are randomly selected for each split in each tree. This distinguishes random forests from bagging, also known as bootstrap aggregation (see Breiman 1996a, 1996b), in which all predictors are considered for each

split. In random forests, m predictors are chosen as split candidates from a set of p variables, i.e., the variable on which the cut is made is selected from m predictors. In general, $m \approx \sqrt{p}$ is the most common choice. In the bagging method, $m = p$, and the strongest predictor for most or all trees is used to make the first split at the top of the tree. This will result in a similarity of trees, so the predictions will be highly correlated. Random forests eliminate this. In both random forests and bagging, the average of the predictions from the regression trees is the prediction. The model of random forests can be expressed by the following formula (see James et al. 2013):

$$\hat{f}_{avg}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{b}(x). \tag{26}$$

For this prediction, the bias is low, and the variance is reduced by averaging a number B of trees (more so for random forests than for bagging). You can use hundreds or thousands of trees to make a single prediction. The optimal number of trees can be optimized through a validation process. In addition to the number of trees, other hyperparameters can be optimized during the process, such as the number of variable candidates at each split or the maximum number of leaves.

The gradient boosting method was developed by Friedman (2001) with the aim of transforming weak learning models (decision trees) into strong learners through a sequential procedure where at each iteration a new decision tree is built based on the previous trees. Thus, the decision trees are not independent. The boosting method is also a combination of a large number of decision trees. For the current model, a decision tree is fitted to the residuals of the previous weak learner model. This decision tree is taken into account for the fitted function, i.e., the residuals are updated, and so on. Trees on residuals tend to be smaller with a small number of terminal nodes. The shrinkage parameter λ can further slow down the process. This is called the learning rate, which is a small positive number. Typical values are 0.01 or 0.001. If λ is very small, a large number of trees will be needed. If each tree is built on a random subsample of the training data, this is called stochastic gradient boosting. The output of the gradient boosting model is as follows (see James et al. 2013):

$$\hat{f}(x) = \sum_{b=1}^{B}\lambda\,\hat{f}^{b}(x). \tag{27}$$

For gradient boosting, in addition to the number of trees and the learning rate, tuning parameters such as the subsampling rate or the interaction depth of a tree can also be optimized. The subsampling rate determines the proportion of the training data that is used to grow the trees. If the interaction depth is one ($d = 1$), then there is a single split. Such boosting models often work well.

II. Research review

This chapter provides an overview of the data and methodology used in this research. The first subsection is *about the data* used in the research. This is followed by a description of *the maximum likelihood estimation method* and the methodological considerations and metrics relevant to the four research topics. *Choosing the length of the base period for modeling* is important for Hungarian mortality data, given that improvements in mortality have not been uniform over the past few decades. *Calculating age-standardized mortality rates* provides advance information on trends in mortality change without the need to fit a mortality model. It is important to understand the *relationship between central and initial mortality rates* in order to switch between mortality rates predicted using models fitted under Poisson and Binomial distribution assumptions. *Measuring the similarity of two mortality indices* may be necessary if we want to check in advance which of the mortality index(es) in a multi-population model should be a common parameter(s). *The Box–Jenkins model selection technique* and the analysis of *outliers in the time series* are both important for predicting time-dependent mortality indices.

For all four research topics, we fitted several mortality models, including multi-population models. In each case, we compared *the goodness of fit* and predictive accuracy of the models to select the best performing model. Therefore, when comparing different mortality models, an analysis of the goodness of fit is inevitable. This can be referred to as an in-sample and an out-of-sample analysis. The latter requires a test period that is different from the base period. This chapter concludes with a discussion of these metrics.

About the data

For the doctoral research, the number of deaths and the population by age and sex were needed to fit stochastic mortality models. In addition, the data were disaggregated by region and by cause of death for two research topics. The data were provided by the

HCSO. However, the necessary data, disaggregated by sex and age, are also available in the Human Mortality Database (HMD), and not only for Hungary.[9] The data for HMD is also provided by HCSO. This database contains annual age-specific Hungarian mortality data from 1950 onwards and population data in the same structure. The longest possible time series is preferred, but data from the 1950s are not used in the doctoral research. The reason for this is that data for the population aged 86–90 are only available by age year from 1960 onwards. In the 1950s, the population aged 85 and over was treated as an aggregate group.

In Chapter IV.1, we analyzed data at regional level, which are only available from 1970 onwards. We were also able to consider data by cause of death from 1970 onwards (see Chapter IV.2). Converted age-specific data by cause of death are available in the database from 1970 onwards. However, cause-specific mortality data by age group are also available in the Population of Hungary and Demographic Yearbook (HCSO 1955–1964 and 1965–present) publications from 1955 onwards. The first and last observed years for which data were used to fit the models for the regional and cause-specific projections were 1970 and 2021, and for the rotation and machine learning research topics were 1960 and 2022 (see Chapters IV.3 and IV.4). The HCSO follows the International Classification of Diseases (ICD) guidelines for classifying causes of death. Different versions of the ICD (from ICD-8 to ICD-10) were in use between 1970 and 2021. The converted data means that mortality by cause of death is comparable over time.[10]

In this study, all models were estimated by age in years, but the mortality data for the age group 90 and over were aggregated. The reason for this is that the population data are available in this structure until 2011,[11] and the uncertainty of the estimation may be increased for smaller age groups above 90 years. Various smoothing methods can be used to estimate age-specific mortality rates for the oldest persons (see Thatcher et al. 1998). The estimation of mortality rates for people aged 90 and over was not included in the doctoral research. However, a smoothing method is in use for the oldest age groups in mortality projections in the HCSO.

Age-specific population data for the 1980s and 1990s have been adjusted in this research. We used the revised total population figures from the 1990 and 2001 censuses

[9] Available at https://www.mortality.org/. Data that can be downloaded from this database do not include the number of deaths in the unknown age.
[10] The ICD-9 to ICD-10 conversion table can be found on the HCSO website: https://www.ksh.hu/classification_bno_x.
[11] The number of people aged 90 and over is available by age year since 2012.

for the adjustment (see HCSO 1990 and 2001). However, the age breakdown of these figures was not calculated at the time. The age distribution of the nonadjusted population, estimated on the basis of vital events and migration between the two censuses, was used to calculate the age distribution of the revised adjusted total population. Revised population data for the 2010s is now available in the HCSO database. However, we have not used this data for this research. As a further adjustment, the number of deaths of unknown age was not ignored in the analysis. We divided the deaths by the proportion of deaths of known age, and added the unknowns to the deaths of known age. This method is consistent with that used by HMD (see Wilmoth et al. 2021). After distributing the unknowns, the number of deaths was rounded to the nearest integer. In the following, we introduce the maximum likelihood estimation method used, and describe considerations related to the current data of the base period, the forecasting process and the comparison of the accuracy of the mortality models.

The maximum likelihood estimation method

In this study, stochastic mortality models were fitted using maximum likelihood estimation with Newton–Raphson iteration method. All models were numerically programmed in R software (R Core Team 2021), and we did not use any packages developed by others. The iteration continued until the increase in the log-likelihood function was less than 10^{-6}. In the following, we present the iterative updating scheme in the context of two mortality models, the Poisson and the Binomial Lee–Carter models. For more models and details, see the online Appendix here and here.

We note that of the above models, the Lee–Carter model with a Poisson or Binomial distribution, the CBD and the full Plat model can be fitted using the *StMoMo* package in R (see Villegas et al. 2018). The Lee–Carter model, the coherent model of Li–Lee (2005) and the rotated Lee–Carter model (see Li et al. 2013 and Chapter I.2.4) can all be applied using the *MortCast* package.[12] The *demography* package[13] can also be used to fit the Lee–Carter model and its Lee–Miller version. The Lee–Carter model is also available with the *LifeMetrics* open source R code[14] and the *ilc* package (see Butt et al. 2014). These two packages can be used for many other models, such as mortality models

[12] See https://CRAN.R-project.org/package=MortCast.
[13] See https://CRAN.R-project.org/package=demography.
[14] See https://www.macs.hw.ac.uk/~andrewc/lifemetrics/.

with a cohort effect (e.g., the age–period–cohort model, see Currie 2006). The CBD and full Plat models can also be fitted using the *LifeMetrics* code. Therefore, only some of the models we are investigating can be fitted using a developed R package. All other mortality models we use require numerical programming to fit. Overall, we applied our own programming procedure to all mortality models to ensure comparability.

The log-likelihood function of the **Poisson Lee–Carter model** can be formulated as follows:

$\ell_j(\theta) = \sum_{xt}\{D_{xtj}\,(\alpha_{xj} + \beta_{1,xj}\kappa_{1,tj}) - E^c_{xtj}e^{\alpha_{xj}+\beta_{1,xj}\kappa_{1,tj}}\}$ + constant

$\theta = (\alpha_{xj}, \beta_{1,xj}, \kappa_{1,tj})$

The parameters θ are updated using the following iterative scheme:

$\widehat{\theta}^{(v+1)} = \widehat{\theta}^{(v)} - \frac{\partial \ell_j{}^{(v)}/\partial\theta}{\partial^2 \ell_j{}^{(v)}/\partial\theta^2}$ where $\ell_j{}^{(v)} = \ell_j{}^{(v)}(\widehat{\theta}^{(v)})$

<u>The iterative updating scheme</u>

1. *Get initial values for the parameters*

$\hat{\alpha}_{xj} = \frac{1}{n_Y}\sum_t ln\frac{D_{xtj}}{E^c_{xtj}}$ where n_Y is the total number of calendar years

$\hat{\kappa}_{1,tj} = (n_Y, \dots, 1) - \frac{(n_Y+1)}{2}$ where n_Y is the total number of calendar years

$\hat{\beta}_{1,xj} = \frac{1}{n_A}$ where n_A is the total number of ages

2. *Update the parameter* $\hat{\alpha}_{xj}$

$$\hat{\alpha}^{(v+1)}_{xj} = \hat{\alpha}^{(v)}_{xj} + \frac{\sum_t \left(D_{xtj} - E^c_{xtj}e^{\hat{\alpha}^{(v)}_{xj}+\hat{\beta}^{(v)}_{1,xj}\hat{\kappa}^{(v)}_{1,tj}}\right)}{\sum_t \left(E^c_{xtj}e^{\hat{\alpha}^{(v)}_{xj}+\hat{\beta}^{(v)}_{1,xj}\hat{\kappa}^{(v)}_{1,tj}}\right)}$$

Adjust the parameter $\hat{\alpha}_{xj} \rightarrow \hat{\alpha}_{xj} = \hat{\alpha}_{xj} + \hat{\beta}_{1,xj}\frac{1}{n_Y}\sum_t \hat{\kappa}_{1,tj}$

Calculate the log-likelihood function ℓ_j

3. *Update the parameter* $\hat{\kappa}_{1,tj}$

$$\hat{\kappa}^{(v+1)}_{1,tj} = \hat{\kappa}^{(v)}_{1,tj} + \frac{\sum_x \left(\left(D_{xtj} - E^c_{xtj}e^{\hat{\alpha}^{(v)}_{xj}+\hat{\beta}^{(v)}_{1,xj}\hat{\kappa}^{(v)}_{1,tj}}\right)\hat{\beta}^{(v)}_{1,xj}\right)}{\sum_x \left(E^c_{xtj}e^{\hat{\alpha}^{(v)}_{xj}+\hat{\beta}^{(v)}_{1,xj}\hat{\kappa}^{(v)}_{1,tj}}\left(\hat{\beta}^{(v)}_{1,xj}\right)^2\right)}$$

Adjust the parameter $\hat{\kappa}_{1,tj}$ in order to satisfy the constraint →

$\hat{\kappa}_{1,tj} = \sum_x \hat{\beta}_{1,xj}\,(\hat{\kappa}_{1,tj} - \frac{1}{n_Y}\sum_t \hat{\kappa}_{1,tj})$

Recalculate the log-likelihood function: compute $\ell_j^{updated}$

4. *Update the parameter* $\hat{\beta}_{1,xj}$

$$\hat{\beta}_{1,xj}^{(v+1)} = \hat{\beta}_{1,xj}^{(v)} + \frac{\sum_t \left(\left(D_{xtj} - E_{xtj}^c e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} \right) \hat{\kappa}_{1,tj}^{(v)} \right)}{\sum_t \left(E_{xtj}^c e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} \left(\hat{\kappa}_{1,tj}^{(v)} \right)^2 \right)}$$

Adjust the parameter $\hat{\beta}_{1,xj}$ in order to satisfy the constraint $\rightarrow \hat{\beta}_{1,xj} = \hat{\beta}_{1,xj} \frac{1}{\sum_x \hat{\beta}_{1,xj}}$

Recalculate the log-likelihood function: compute $\ell_j^{updated}$

5. *Check convergence*: the difference between the log-likelihood and the updated log-likelihood is less than 10^{-6}

$\Delta \ell_j = \ell_j - \ell_j^{updated}$ ($\ell_j^{updated}$ is the updated log-likelihood from step 4)
In case of nonconvergence ($\Delta \ell_j > 10^{-6}$), select different initial values for the parameters.

The log-likelihood function of the **Binomial Lee–Carter model** can be formulated as follows:

$\ell_j(\theta) = \sum_{xt} \{ D_{xtj} ln\, q_{xtj} + (E_{xtj}^0 - D_{xtj}) ln(1 - q_{xtj}) \}$ + constant
where $q_{xtj} = e^{\alpha_{xj} + \beta_{1,xj} \kappa_{1,tj}} / (1 + e^{\alpha_{xj} + \beta_{1,xj} \kappa_{1,tj}})$
$\theta = (\alpha_{xj}, \beta_{1,xj}, \kappa_{1,tj})$
The parameters θ are updated using the following iterative scheme:
$\hat{\theta}^{(v+1)} = \hat{\theta}^{(v)} - \frac{\partial \ell_j{}^{(v)} / \partial \theta}{\partial^2 \ell_j{}^{(v)} / \partial \theta^2}$ where $\ell_j{}^{(v)} = \ell_j{}^{(v)}(\hat{\theta}^{(v)})$

<u>The iterative updating scheme</u>

1. *Get initial values for the parameters*

$\hat{\alpha}_{xj} = \frac{1}{n_Y} \sum_t ln \frac{D_{xtj}}{E_{xtj}^0}$ where n_Y is the total number of calendar years
$\hat{\kappa}_{1,tj} = (n_Y, \ldots, 1) - \frac{(n_Y + 1)}{2}$ where n_Y is the total number of calendar years
$\hat{\beta}_{1,xj} = \frac{1}{n_A}$ where n_A is the total number of ages

2. *Update the parameter* $\hat{\alpha}_{xj}$

$$\hat{\alpha}_{xj}^{(v+1)} =$$
$$= \hat{\alpha}_{xj}^{(v)} + \frac{\sum_t \left((D_{xtj} e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} - E_{xtj}^0 e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} + D_{xtj}) / (e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} + 1) \right)}{\sum_t \left((E_{xtj}^0 e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}}) / (e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} + 1)^2 \right)}$$

Adjust the parameter $\hat{\alpha}_{xj} \rightarrow \hat{\alpha}_{xj} = \hat{\alpha}_{xj} + \hat{\beta}_{1,xj} \frac{1}{n_Y} \sum_t \hat{\kappa}_{1,tj}$
Calculate the log-likelihood function ℓ_j

3. *Update the parameter* $\hat{\kappa}_{1,tj}$

$$\hat{\kappa}_{1,tj}^{(v+1)} =$$

$$= \hat{\kappa}_{1,tj}^{(v)} + \frac{\sum_x \left(\left((D_{xtj} e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} - E_{xtj}^{0} e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} + D_{xtj}) / (e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} + 1) \right) \hat{\beta}_{1,xj}^{(v)} \right)}{\sum_x \left(\left((E_{xtj}^{0} e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}}) / (e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} + 1)^2 \right) \left(\hat{\beta}_{1,xj}^{(v)} \right)^2 \right)}$$

Adjust the parameter $\hat{\kappa}_{1,tj}$ in order to satisfy the constraint →

$$\hat{\kappa}_{1,tj} = \sum_x \hat{\beta}_{1,xj} \left(\hat{\kappa}_{1,tj} - \frac{1}{n_Y} \sum_t \hat{\kappa}_{1,tj} \right)$$

Recalculate the log-likelihood function: compute $\ell_j^{updated}$

4. *Update the parameter* $\hat{\beta}_{1,xj}$

$$\hat{\beta}_{1,xj}^{(v+1)} =$$

$$= \hat{\beta}_{1,xj}^{(v)} + \frac{\sum_t \left(\left((D_{xtj} e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} - E_{xtj}^{0} e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} + D_{xtj}) / (e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} + 1) \right) \hat{\kappa}_{1,tj}^{(v)} \right)}{\sum_t \left(\left((E_{xtj}^{0} e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}}) / (e^{\hat{\alpha}_{xj}^{(v)} + \hat{\beta}_{1,xj}^{(v)} \hat{\kappa}_{1,tj}^{(v)}} + 1)^2 \right) \left(\hat{\kappa}_{1,tj}^{(v)} \right)^2 \right)}$$

Adjust the parameter $\hat{\beta}_{1,xj}$ in order to satisfy the constraint → $\hat{\beta}_{1,xj} = \hat{\beta}_{1,xj} \frac{1}{\sum_x \hat{\beta}_{1,xj}}$

Recalculate the log-likelihood function: compute $\ell_j^{updated}$

5. *Check convergence*: the difference between the log-likelihood and the updated log-likelihood is less than 10^{-6}

$\Delta \ell_j = \ell_j - \ell_j^{updated}$ ($\ell_j^{updated}$ is the updated log-likelihood from step 4)

In case of nonconvergence ($\Delta \ell_j > 10^{-6}$), select different initial values for the parameters.

Choosing the length of the base period for modeling

According to Lee–Carter (1992), the forecast is not sensitive to shortening the base period, e.g., from 90 to 30 years, but the period of 10–20 years is already too short, and leads to instability. Therefore, sufficiently long time series of deaths and population should be chosen. By estimating over a longer and a shorter period, the similarity of the results means that the structural homogeneity of the chosen period is achieved. There is a possible test (out-of-sample analysis) of the mortality model if we make the forecast for

a shorter period for which the data are already known. In this way, we can compare the actual data with the estimated values.

The choice of the length of the base period may be influenced by the presence of a structural change in the time series. The presence of the structural change may affect the forecast of the mortality index(es). In Hungary, the trend of improving mortality started from a high level of mortality in the 1990s. The mortality crisis started in the mid-1960s and continued until the mid-1990s. In addition to political and economic changes, mortality was less favourable in the 1990s, especially for men. Baran et al. (2007) fitted different mortality models to Hungarian data, and tried to find the ideal length of the base period due to the impact of the political and economic regime change in Hungary in 1989–1990. Bajkó et al. (2015) also tested the Lee–Carter model on base periods of different lengths. According to their results, a time series starting after 1980 may be the most suitable for preparing forecasts. The choice of base period length has an impact on how long a forecast can be made. For all four research themes, data from 1970 to 2021 or 1960 to 2022 were used to produce projections to 2050. So in all cases the base period was longer than the projection period.

Calculating age-standardized mortality rates

Age-standardized mortality rates (ASMRs) can be used to analyze mortality trends based on time series data. The ASMR is an aggregated measure calculated from crude death rates. After dividing the number of deaths (D_{xtj}) by the population (E_{xtj}), the quotient is multiplied by the age-varying weights of the standard population. According to Wen et al. (2021), the formula for the ASMR in calendar year t for subpopulation j can be written as follows:

$$ASMR_{tj} = \sum_{x\in\chi} \frac{D_{xtj}}{E^{c}_{xtj}} w_x, \qquad w_x = \frac{E^{c,s}_{x}}{\sum_{x\in\chi} E^{c,s}_{x}} \tag{28}$$

where the weights w_x denote the age-specific proportions of the standard population $E^{c,s}_{x}$. In this study, we chose the mid-year population (central exposure) to calculate the ASMRs. In addition, the data from the first year of the base period (1960 or 1970) was always chosen as the standard population. Age range χ denotes age in years, ranging from 0 to 90 years for the regional analysis of mortality and from 60 to 90 years for the analysis

by cause of death. ASMRs were defined separately by subpopulation and sex. It is advisable to calculate ASMRs before fitting stochastic mortality models because the slope of the mortality index, which describes the long-term trend in mortality change, and the slope of the weighted aggregate index are very similar.

Relationship between central and initial mortality rates

In this study, we fitted some mortality models that assume a Binomial distribution for the number of deaths, and use a logit link function. For these models, we used initial exposure. For the Poisson distribution assumption and log link function, central exposure is used. Central mortality rates can be calculated from the initial mortality rates, making it easier to compare results from different models. The initial mortality rates can be converted to central mortality rates using the following relationship:

$$m_{xtj} = \frac{q_{xtj}}{1-0.5\ q_{xtj}}, \tag{29}$$

or which is almost identical to:

$$m_{xtj} = -ln(1 - q_{xtj}). \tag{30}$$

Equation 29 assumes that those, who died in year t lived, on average, until the middle of the period. Equation 30 assumes that the force of mortality is constant over age x and in year t.

Measuring the similarity of two mortalitiy indices

The mortality indices of the selected single-population models, which describe trends in mortality change, were clustered to create multi-population models in the analysis by cause of death. Before clustering, the similarity of the mortality indices and the distance between them had to be measured. The dynamic time warping (DTW) method was used for this measurement. The R packages *dtw* and *dtwclust*, developed by Giorgino (2009) and Sardá-Espinosa (2017, 2019), are suitable for investigating the optimal alignment between time series and for clustering time series. Bellman–Kalaba (1959) invented the

DTW method, which is a widely used technique for measuring the similarity between two temporal sequences. The authors originally applied this method for speech recognition.

Considering two cause-of-death-specific mortality indices of a mortality model, we present the steps to measure the similarity between two time series. Let $\kappa_{j,h}^{(i)}$ be the h-th element of index i for cause of death j, and let $\kappa_{-j,v}^{(i)}$ be the v-th element of index i for another cause of death. This applies for i =1, 2, 3, and j = 1, …, 9. The minus sign before j indicates that it is another death, not the j-th. The DTW method finds the optimal alignment between time series at the lowest possible cost. The first and last elements of the series are matched, while other elements can be warped in time. The local cost between the h-th and v-th elements of $\kappa_j^{(i)}$ and $\kappa_{-j}^{(i)}$ can be measured as follows:

$$c\left(\kappa_{j,h}^{(i)}, \kappa_{-j,v}^{(i)}\right) = \left|\kappa_{j,h}^{(i)} - \kappa_{-j,v}^{(i)}\right|. \tag{31}$$

The local cost matrix (LCM) is the result of calculating the local cost for each pair of elements in the time series. In the comparison, the cause-of-death-specific mortality indices originate from the same mortality model, with the same superscript. In this study, the length of the mortality indices is the same, so the LCM has $t \times t$ dimensions. An LCM is created for each comparison of two time series. The l_p norm[15] determined between $\kappa_{j,h}^{(i)}$ and $\kappa_{-j,v}^{(i)}$ for each element (h, v) is used to calculate the LCM. See Sardá-Espinosa (2017, 2019).

$$LCM(h, v) = \left(\sum_i \left|\kappa_{j,h}^{(i)} - \kappa_{-j,v}^{(i)}\right|^p\right)^{1/p} \tag{32}$$

The DTW algorithm moves step by step through the matrix from the LCM(1,1) position to the LCM(t,t) position. This warping path minimizes the alignment between the time series, and the costs are then aggregated. The final distance value is calculated as follows (see Sardá-Espinosa 2017):

[15] The two most commonly used norms, l_1 and l_2, i.e., the Manhattan and Euclidean distances, have limitations, For example, they are only applicable to time series of equal length, as described by Sardá-Espinosa (2017, 2019). In this study, p was set to 2 and the length of the mortality indices was the same, so this was not a problem.

$$DTW_p(\kappa_j, \kappa_{-j}) = \left(\sum \frac{m_\phi LCM(k)^p}{M_\phi}\right)^{1/p}, \forall k \in \phi \quad (33)$$

where $\phi = \{(1,1), \ldots, (t,t)\}$, m_ϕ is the per-step weighting coefficient, and M_ϕ denotes the corresponding normalization constant (see Giorgino 2009). With respect to DTW_p, p refers to the l_p norm.

The Box–Jenkins model selection technique

One of the main objectives of mortality models is to predict mortality rates by age. This requires time series analysis of the mortality index(es) and the cohort effect. In general, the time series of the observed data should be longer than the number of years of the projection. An important feature of the projections is that the standard errors become larger, and the prediction intervals wider over time. Assuming the independence of the mortality index(es) and the cohort effect, we can separately examine the best ARIMA models for forecasting (see, e.g., Plat 2009, Villegas et al. 2018).

According to the Box–Jenkins procedure (see Enders 2014), the selection of the appropriate time series model consists of three steps. The first is identification where we examine the time series to identify, for example, outliers, missing values, and structural breaks in the data. The presence of a unit root can be tested using the augmented Dickey–Fuller (ADF) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. We plot the autocorrelation function (ACF) and the partial autocorrelation function (PACF) to estimate how many lagged values are needed for the best ARIMA model. In the second step, we fit a parsimonious model because they are better predictors than those with too many parameters. The third step is to check that the residuals of the estimated model definitely follow a white noise process. For example, the Breusch–Godfrey test and the Ljung–Box Q-statistic can be used to test residuals for autocorrelation, while the Shapiro–Wilk test can be used to check for normal distribution. In this study, we followed the Box–Jenkins model selection technique to predict the mortality index(es).

Outliers in the time series

An important issue in fitting mortality models can be the detection of outliers. The study by Lee–Carter (1992) dealt with the effect of the Spanish flu on mortality, and this aspect

is very relevant in recent years because of the COVID-19 epidemic. In their analysis, the effect of the Spanish flu epidemic of 1918 was filtered out in an additive way using a dummy variable: a binary variable was included in the equation predicting the mortality index to indicate the year in which mortality was particularly high. Because of the impact of the COVID-19 epidemic, it is practical to follow a similar approach when using data from recent years to predict mortality. During the doctoral research, we found that for the Hungarian data, the data for 2020 and 2021 should be treated as outliers. If we do not filter for outliers, then we are looking at the epidemic as an event that will reappear in the future. This only affects the prediction interval, not the prediction itself. If no dummy variable is used, the prediction interval will be wider. This binary variable can also be interpreted as a time series. Box plots can be used to identify the value and location of the outliers in the undifferentiated time series of the stochastic parameters.

The goodness of fit

There are several methods that can be used to facilitate the choice between the selected mortality models. One method is to compare the estimated mortality rates with the actual values. If we are making a forecast for a period for which we already know the data, we can analyze the similarity between the predicted mortality rates and the current values. In this way we can determine which model is more accurate. In the following, we present several possible measures that can be used to compare the results of mortality models.

The choice between models can be supported by comparing the values of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). To compare two models in this respect, they should be estimated for the same time period (see, e.g., Enders 2014). The lower the AIC or BIC value, the better the model. The AIC and BIC values can be calculated as follows (see, e.g., Villegas et al. 2018):

$$AIC = 2k - 2L, \tag{34}$$

$$BIC = k\ ln(n) - 2L \tag{35}$$

where k denotes the number of the parameters, n refers to the number of observation, and L is the maximum log-likelihood value.

The mean absolute percentage error (MAPE) is also an appropriate measure to assess the accuracy of different mortality models (see, e.g., Li 2013). The lowest MAPE value is preferred when comparing models.

$$MAPE_j = \frac{1}{n}\sum_{xt}\left|\frac{\widehat{m}_{xtj}-D_{xtj}/E^c_{xtj}}{D_{xtj}/E^c_{xtj}}\right| \text{ or } MAPE_j = \frac{1}{n}\sum_{xt}\left|\frac{\hat{q}_{xtj}-D_{xtj}/E^0_{xtj}}{D_{xtj}/E^0_{xtj}}\right| \qquad (36\text{ a–b})$$

where n refers to the number of observations, D_{xtj} is the observed number of deaths, E_{xtj} denotes the central or initial exposure, depending on the mortality model, and $\widehat{m}_{xtj}$ and $\hat{q}_{xtj}$ are the fitted central or initial mortality rates at age x year t and subpopulation j, respectively. In this study, two metrics were calculated during hyperparameter optimization for tree-based machine learning methods: mean squared error (MSE) and mean absolute error (MAE) – see, e.g., Schnürch–Korn (2022). Since we applied machine learning algorithms to the residuals of stochastic mortality models, the MSE and MAE can be written as follows:

$$MSE_j = \frac{1}{n}\sum_{xt}(\widehat{res}^{model}_{xtj} - res^{model}_{xtj})^2, \qquad (37)$$

$$MAE_j = \frac{1}{n}\sum_{xt}\left|\widehat{res}^{model}_{xtj} - res^{model}_{xtj}\right| \qquad (38)$$

where res^{model}_{xtj} is the difference between the actual and the fitted log mortality rates, and $\widehat{res}^{model}_{xtj}$ denotes the predicted residuals on a test set using a machine learning method. The lowest MSE and MAE values are preferable.

The explanation ratio (see, e.g., Wen et al. 2021, Li–Lee 2005) is also a possible measure of goodness of fit, which can be calculated as follows:

$$R_j = 1 - \frac{\sum_{xt}(ln\frac{D_{xtj}}{E^c_{xtj}} - ln\,\widehat{m}_{xtj})^2}{\sum_{xt}(ln\frac{D_{xtj}}{E^c_{xtj}} - \alpha_{xj})^2} \text{ or } R_j = 1 - \frac{\sum_{xt}(ln\frac{D_{xtj}}{E^0_{xtj}} - ln\,\hat{q}_{xtj})^2}{\sum_{xt}(ln\frac{D_{xtj}}{E^0_{xtj}} - \alpha_{xj})^2} \qquad (39\text{ a–b})$$

where D_{xtj}, E^c_{xtj}, E^0_{xtj}, $\widehat{m}_{xtj}$, and $\hat{q}_{xtj}$ notations are the same as before. The term α_{xj} is the average of log death rates by age over the years of the base period.

$$\alpha_{xj} = \frac{1}{n_Y}\sum_t ln\frac{D_{xtj}}{E^c_{xtj}} \text{ or } \alpha_{xj} = \frac{1}{n_Y}\sum_t ln\frac{D_{xtj}}{E^0_{xtj}} \quad (40\text{ a–b})$$

where n_Y equals to the total number of calendar years. In this study, the central and initial exposures were used to calculate the explanation ratio. This measure shows that the higher the value for a mortality model, the better the goodness of fit.

Another measure of goodness of fit, the likelihood ratio (LR) statistic, is calculated as follows, according to Brouhns et al. (2002):

$$LR_j = 2\sum_{xt} D_{xtj}\, ln\left(\frac{D_{xtj}}{\widehat{D}_{xtj}}\right). \quad (41)$$

This measure is based on a comparison of the current (D_{xtj}) and estimated ($\widehat{D}_{xtj}$) number of deaths. The lower value is preferred for a model. Another way to compare different mortality models is to analyze the standardized residuals, which can be determined as follows (see, e.g., Wen et al. 2021):

$$Z_{xtj} = \frac{D_{xtj} - E^c_{xtj}\hat{m}_{xtj}}{\sqrt{E^c_{xtj}\hat{m}_{xtj}}} \text{ or } Z_{xtj} = \frac{D_{xtj} - E^0_{xtj}\hat{q}_{xtj}}{\sqrt{E^0_{xtj}\hat{q}_{xtj}}} \quad (42\text{ a–b})$$

where D_{xtj}, E^c_{xtj}, E^0_{xtj}, $\hat{m}_{xtj}$, and $\hat{q}_{xtj}$ notations are the same as above. We can plot the residuals against age, calendar year, and year of birth using heatmaps and scatter plots. Based on these analyses, it can be determined whether the cohort effect needs to be taken into account for a given model. The better the model fit, the more random the residuals become. The characteristic diagonal patterns in the heatmap may suggest that it is worth adding a cohort effect to the model.

Some mortality models can be considered as nested models. The more general model is the more extended model, it has the most parameters. The nested model is part of the general model. According to Plat (2009), the LR test for general and nested models is more appropriate for model selection than information criteria. According to the LR test, the null hypothesis is that the nested model is the correct model against the general model. The formula for the LR test statistic is:

$$\xi_{\text{LR}} = 2(L_{general} - L_{nested}) \quad (43)$$

where $L_{general}$ means the maximum log-likelihood value of the general model, and L_{nested} refers to the log-likelihood of the nested model. ξ_{LR} has a χ^2 distribution with J degrees of freedom under the null hypothesis. J is the number of additional estimated parameters in the general model compared to the nested model. We can reject the null hypothesis if:

$$\xi_{LR} > \chi^2_{J,\alpha} \tag{44}$$

where α refers to the significance level.

These metrics, which measure the goodness of fit, were used to perform in-sample analyses and to facilitate comparisons between different mortality models or tree-based machine learning methods. The MAPE measure was used for the out-of-sample analysis. The next chapter presents the mortality trends in Hungary over the last decades, broken down by sex, region and cause of death.

III. Description of mortality in Hungary

This chapter presents differences in Hungarian mortality by sex, cause of death and region. The analyses of Józan (2008, 2009, 2012) are important references for understanding Hungarian mortality in recent decades. Józan (2008) covered the period after World War II until 2006. Józan reviewed mortality data from these decades using descriptive statistics, and distinguished three phases of epidemiological development in Hungary. In this doctoral thesis, we used mortality data and time series starting from 1960 or 1970, and we focused mainly on the last two of the three phases distinguished by Józan (2008). According to Józan (2008), the epidemiological development in Hungary has been discontinuous, in contrast to the Western development, with the period between 1948 and 1966 being the 1st phase, referred to as the hopeful beginning. In particular, infant and child mortality fell during this phase, and life expectancy for young adults increased. The period was particularly marked by successes in curing pneumonia, tuberculosis and other infectious diseases.

Józan (2008) found that the years between 1967 and 1993 represented the 2nd phase. This was the period of chronic, qualified epidemiological crisis. The term epidemiological crisis refers to a negative trend in mortality, with life expectancy at birth

falling (or stagnating). It was chronic in the sense that it lasted for decades. The term qualified expresses the fact that this condition affected only a part of the population, mostly middle-aged men, and according to Józan (2008) it was specific to Central and Eastern Europe, and particularly pronounced in Hungary. During this period, the mortality rates of men and women differed: men's mortality rates worsened, while women's mortality rates improved. This is because the loss of life years among middle-aged women was offset by gains in life years among the young and old (see Bálint 2016).

In the 2nd phase, life expectancy declined due to negative trends in neoplasms, cardiovascular and digestive diseases, and deaths from violence.[16] Among diseases of the digestive system, a significant increase in deaths due to alcoholic liver disease should be noted. According to Józan (2008), ischaemic heart disease, alcoholic liver disease, and lung cancer accounted for almost nine-tenths of worsened mortality in the 2nd phase. However, Józan (2008) also noted that between 1967 and 1993, only one third of deaths from chronic ischaemic heart disease were confirmed by autopsy, so its importance may be overestimated. The same is true for liver disease, where about half of the deaths were verified by autopsy. With regard to the increase in mortality from alcoholic liver disease in the early years of the change of political and economic regime, it cannot be excluded that this is partly due to the fact that cirrhosis of the liver, previously classified as a concomitant disease, was diagnosed as the underlying disease in some cases. So this was not a completely real phenomenon.

However, mortality also improved in the 2nd phase for some diseases, such as tuberculosis and malignant neoplasm of the stomach. The latter has been linked to changes in food preservation methods. Lethal diseases and the lifestyle that caused them were already widespread in the 1950s. This led to a negative trend in mortality in the 1960s and 1970s. According to Józan (2008), the increase in slowly progressive diseases in the early 1990s was linked to the insecurity of life following the change of political and economic regime, and an increase in violent mortality. The slowly progressive diseases were mainly due to years and decades of unhealthy lifestyles (e.g., increased alcohol consumption and smoking). This unhealthy lifestyle before the change of political and economic regime was an important cause of the unfavourable mortality in Hungary in the early 1990s. According to Józan (2008), this was exacerbated by the changing

[16] With regard to mortality in Hungary, it may be particularly important to examine the number of suicides within the main group of external causes of morbidity and mortality. However, this study focuses in detail on the main groups of causes of death. See Zonda et al. (2013) for a detailed analysis.

practice of coding deaths in cases such as alcoholic liver disease. The epidemiological crisis reached its peak in 1993.

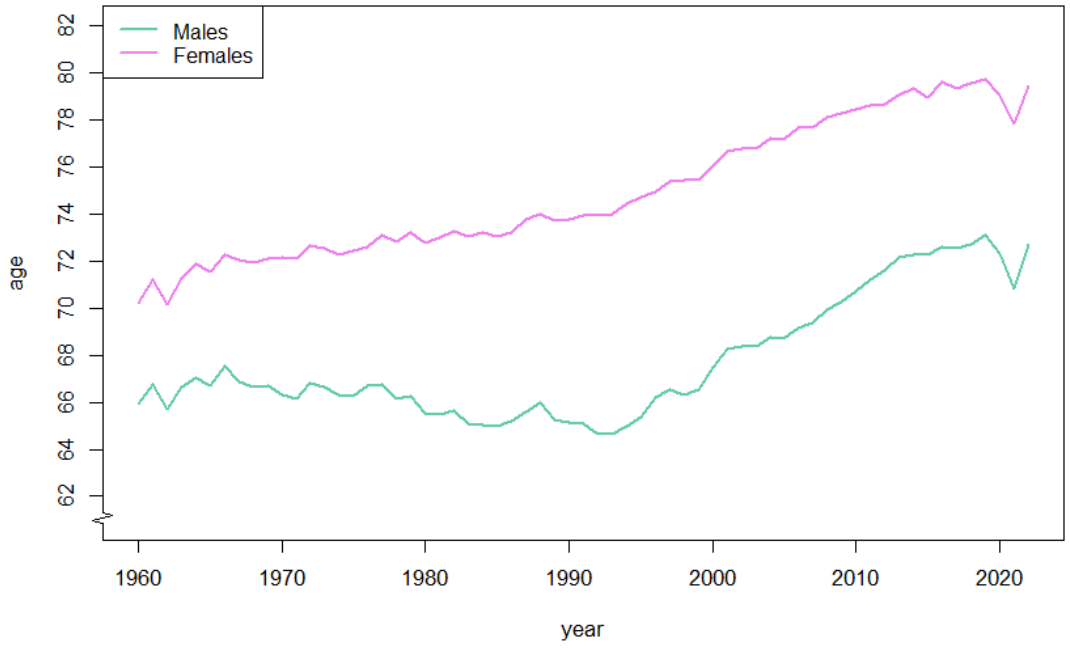
The 3rd phase began in 1994, and was the period of renewal. The turnaround and improvement in mortality coincided with the years following the political and economic regime change. By then, the rate of improvement in child mortality had slowed down, and the increase in life expectancy was more attributable to the middle-aged and elderly. After 1994, men contributed much more than women to the increase in life expectancy. One explanation for this, according to Józan (2008), may be that men started from a very low level in terms of life expectancy, and had a lot of catching up to do. On the other hand, women's life expectancy has risen more slowly because their lifestyles have become more similar to men's, for example, in terms of the prevalence of smoking (see Józan 2006). The study of Hrzic et al. (2020) revealed that European Union Member States with initially higher mortality rates experienced improvements more quickly than those with more favourable initial conditions. Jasilionis et al. (2023) found that the collapse of the communist regimes had an immediate positive effect on population health trends in Central and Eastern European countries.

According to Józan (2008), hypertension, cerebrovascular diseases, cardiovascular diseases, diabetes mellitus, and certain malignant neoplasms of lymphoid and haematopoietic were effectively treated in the 3rd period. This phase of epidemiological development is referred to in the international literature as the cardiovascular revolution (see Meslé–Vallin 2002, Vallin–Meslé 2004, 2009), which reached Hungary after a delay of several decades. Aburto–van Raalte (2018) analyzed mortality in Central and Eastern European countries, including Hungary. Their findings confirmed that the decline in cardiovascular mortality is a delayed phenomenon. Fihel–Pechholdová (2017) also found that, by the 1970s, Western European countries had already achieved success in terms of comprehensive health programs and prevention measures, particularly with regard to cardiovascular mortality. In contrast, health policies in Central and Eastern Europe were not yet effective at that time. Chronic noninfectious diseases predominated in the 3rd phase in Hungary, but these tended to occur at an older age, and disease progression improved (see Józan 2009). The improvement in mortality was mainly due to the spread of health awareness, effective medicine, medical technology and emergency care (see Józan 2012). Epidemiological developments in the 21st century are expected to improve mortality, particularly in old age.

The above three phases clearly show differences in mortality by age, sex and cause of death. Figure 1 confirms the differences in mortality between men and women, illustrating life expectancy at birth[17] by sex in Hungary over recent decades. The difference in life expectancy between men and women was favourable in the early 1960s, when men's life expectancy at birth was only 4–5 years lower than that of women. Thereafter, the longevity gap gradually widened, and was at its widest between 1992 and 1995, when men's life expectancy at birth was 9.3 to 9.5 years less than that of women. Since the second half of the 1990s, men have overcome their considerable disadvantage. Today, the difference in life expectancy at birth between men and women is around 6–7 years, so despite improvements, the favourable levels of the early 1960s have not been reached. Between 1960 and 2022, life expectancy at birth increased by years: it was 65.9 years for men and 70.2 years for women in 1960. In 2022, these values were 72.7 and 79.5 years, respectively. Men and women have followed different development paths. Life expectancy at birth for men decreased between 1960 and 1993, and was lowest in 1992–1993 (64.6 years). For women, the long-term trend has been a gradual increase in life expectancy. In the years 2020 to 2022, life expectancy was lower due to the COVID-19 epidemic.

[17] In calculating the life tables, we mainly followed the methodology of Chiang (1968) and Eurostat (see https://ec.europa.eu/eurostat/cache/metadata/Annexes/demo_mor_esms_an_1.pdf). The average fraction of the year lived by an infant was 0.2, and the fraction of the last year of life was 0.5 for all other cohorts. In addition, the aggregate group of people aged 90 and over was considered as the oldest cohort.

Figure 1. Life expectancy at birth for males and females between 1960 and 2022

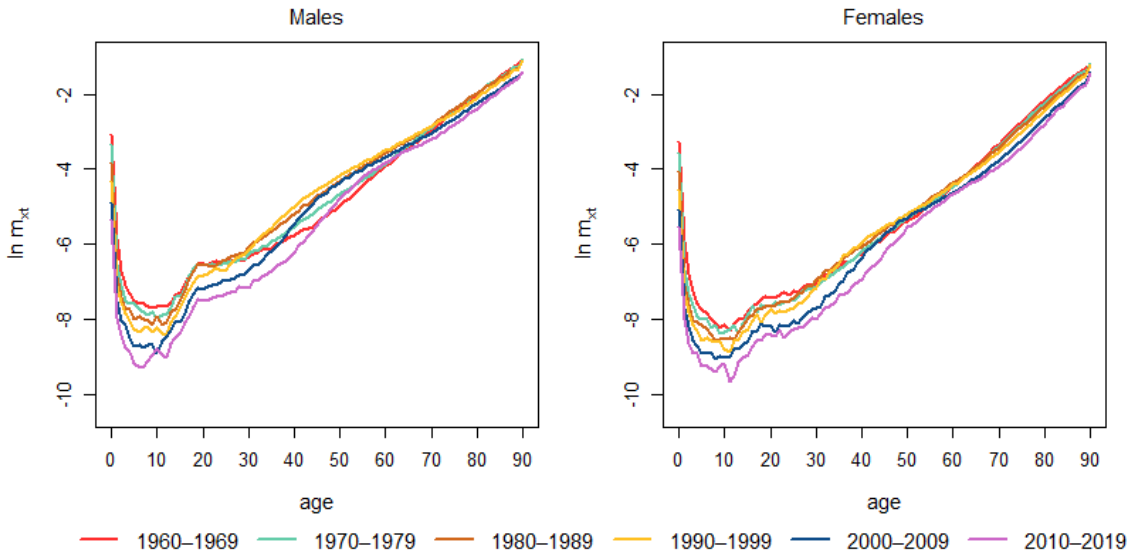


Source: own calculation

Figure 2 shows the logarithms of the smoothed mortality rates. The logarithms of mortality rates are averaged over 10 years, so smoothed values are shown instead of annual log mortality rates. Data for the years 2020–2022 have been excluded from the smoothing due to the COVID-19 epidemic. There was an improvement in mortality during the period under review, with the curves shifting downwards. The difference in mortality rates between men and women is also evident: women have lower log mortality rates than men. Overall, the smallest improvement is observed among people aged 55–65. However, for those aged under 55, the improvement in mortality has been significant, with much lower rates for both men and women between 2010 and 2019. The mortality gap between men and women is particularly wide for those aged around 20. In general, the mortality of women is more favourable, especially in the cohort of 20-year-olds. Men have a lower life expectancy than women at this age, mainly due to the newly obtained driving licence and traffic accidents.[18]

[18] See, e.g., the Tables 6.2.20, 6.2.21 and 6.2.22 in HCSO (2023).

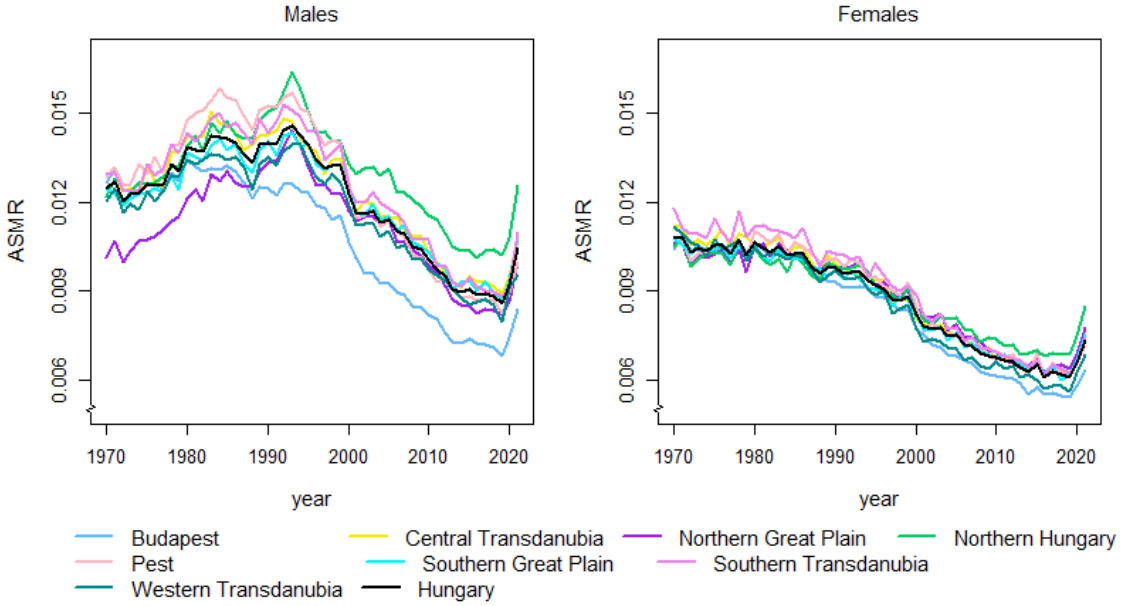
Figure 2. The logarithms of the smoothed central mortality rates between 1960 and 2019



Source: own calculation

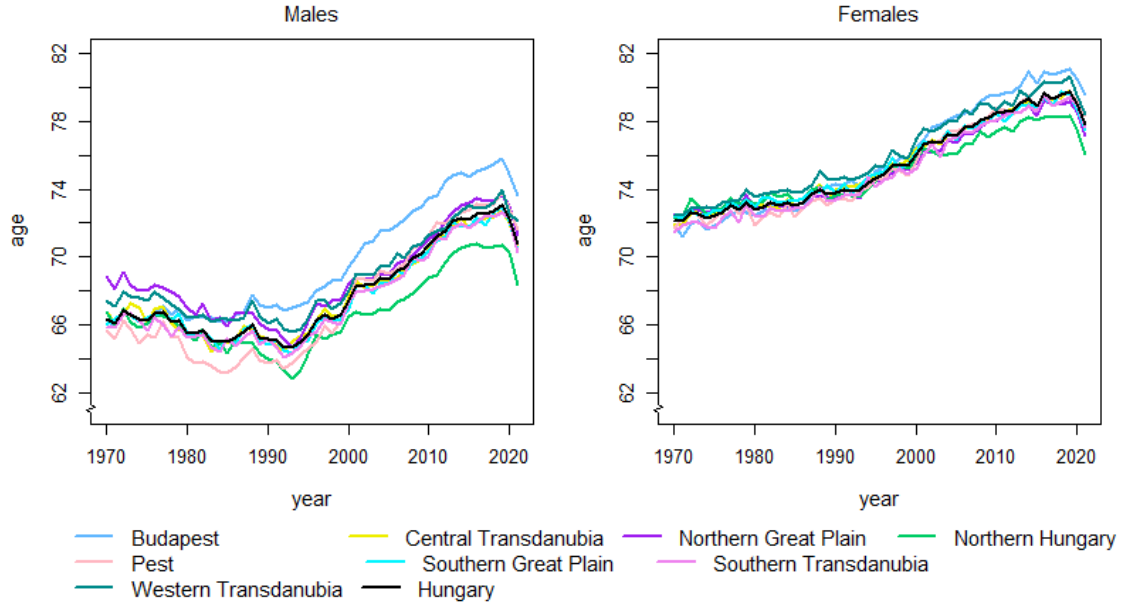
Figures 3–4 show the differences in mortality for men and women by region over the last 50 years. Figure 3 presents the trend in age-standardized mortality rates. The high mortality in the first half of the 1990s was particularly striking in Northern Hungary and the Northern Great Plain. The ASMR values also show that the primary trend in male mortality was downward until the early 1990s. For women, mortality rates have almost stagnated since the 1970s, with a clear improvement since the second half of the 1980s. For both men and women, the improvement in mortality had slowed down in the years before the COVID-19 epidemic. When analyzing life expectancy, Bálint (2011) found that territorial differences were more significant for men than for women. Figures 3–4 also confirm this. Regional differences in both ASMR and life expectancy at birth have been greater for men than for women in recent decades. ASMR values and life expectancy at birth curves are essentially complementary: ASMR values are high when there are many deaths in a given year, and life expectancy is low. In recent decades, life expectancy at birth for both sexes has been the highest in Budapest and the lowest in Northern Hungary.

Figure 3. Age-standardized mortality rates by region between 1970 and 2021



Source: own calculation

Figure 4. Life expectancy at birth by region between 1970 and 2021

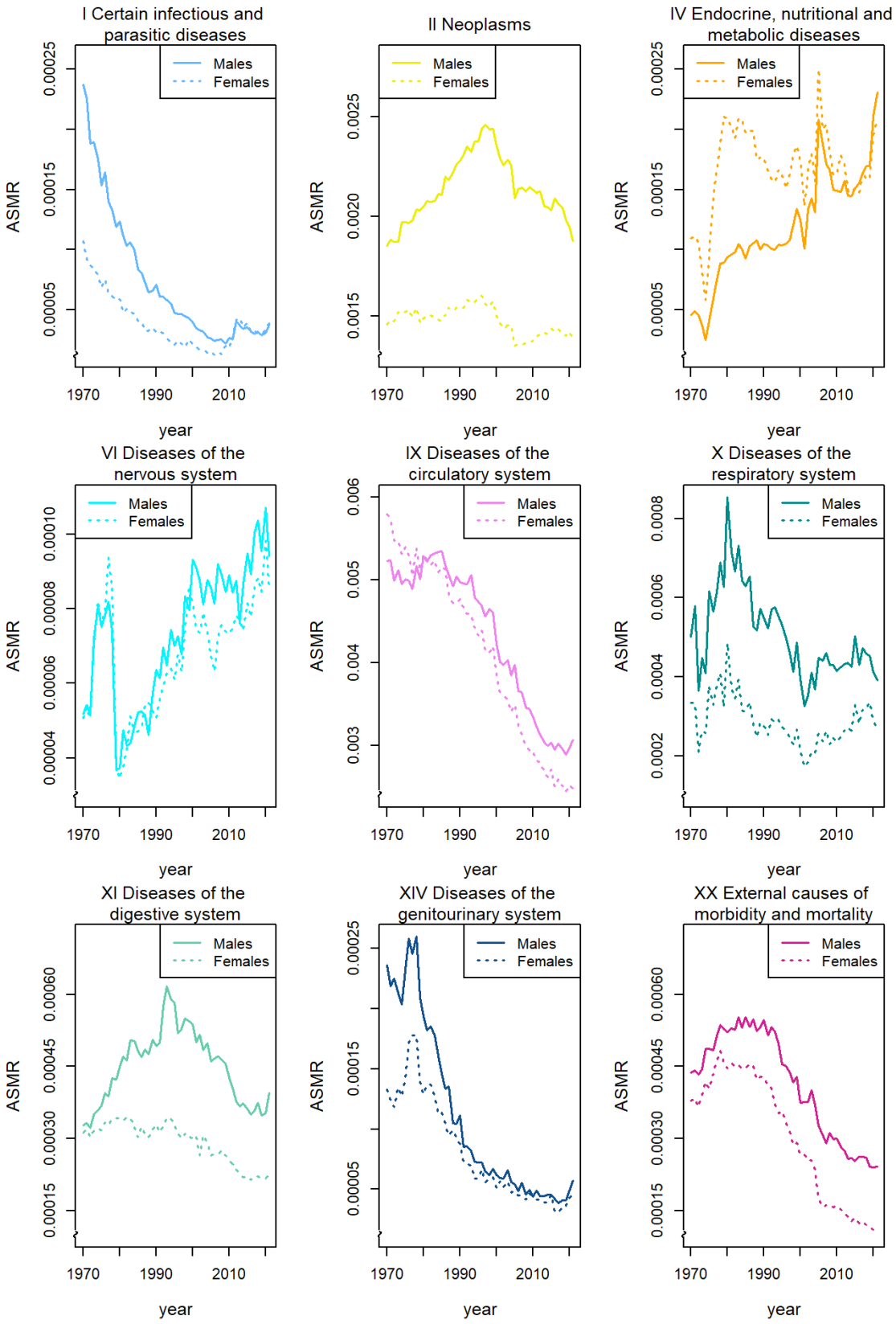


Source: own calculation

Figure 5 visualizes the ASMR values between 1970 and 2021 by nine main causes of death for both sexes separately. The selected causes of death confirm the findings of Józan (2008). A decline in certain infectious and parasitic diseases has been observed in recent decades up to 2010. In the 2nd phase of epidemiological development, neoplasms,

diseases of the digestive system, and external causes of morbidity and mortality have shown a deteriorating trend in mortality, especially for men. A worsening trend was also observed for the main group of diseases of the nervous system, and this is still the case today. In addition, there has been a clear downward trend in endocrine, nutritional and metabolic diseases in recent years.

Figure 5. Age-standardized mortality rates by cause of death between 1970 and 2021



Source: own calculation

IV. Results of modeling mortality in Hungary

In this chapter, we present the results of the four research topics for which we formulated the research questions (see the Introduction chapter). First, we analyze and project regional mortality in Hungary. We then examine mortality by cause of death, followed by addressing rotation. We then examine mortality by cause of death, followed by rotation. Finally, we consider the incorporation of results from tree-based machine learning methods for different standard mortality models. The main structure of each research topic is very similar: an introduction to the topic is followed by a presentation of the fitted model parameters and the in-sample analysis. We then discuss mortality indices and life expectancy forecasts. Next, we present the results of the out-of-sample analysis, and select the mortality model that best fits the test data. Each research topic is concluded with a summary. For some topics, additional analytical issues are raised. The results of the dynamic time warping and clustering methods offer a new perspective on analyzing the causes of death. Investigating the time-dependence of the age-varying parameters is important in the context of rotation, as it helps to understand how these parameters change over time. Selecting the optimal set of hyperparameters is related to machine learning research topics.

IV.1. Fitting and forecasting multi-population mortality models based on Hungarian regional data[19]

Introduction

Multi-population models can be used to model regional mortality data jointly, and attempt to make coherent predictions (see Li–Lee 2005), i.e., that mortality rates do not diverge when compared spatially. For this topic, in addition to the Lee–Carter model,[20] five multi-population models were selected for analysis and prediction. The subpopulations were the Hungarian regions, and the models were fitted separately by sex. We used maximum likelihood estimation, assuming a Poisson distribution for the number of deaths. The base period of observation was 1970 to 2021, and life expectancy was projected to 2050. Data

[19] Varga (2023) is the basis for this chapter.
[20] In this chapter, Lee–Carter model refers to the Poisson version.

were used for the population aged 0–90. The final age category was treated as an aggregated group.

When examining regional mortality rates, the impact of internal migration is worth noting. In this study, the number of deaths was based on temporary rather than permanent residence. Mortality rates were calculated without including people who were homeless, foreign nationals, or individuals whose residence was unknown. The Hungarian population does not live in isolated regions. Consequently, there is migration between regions, which varies in intensity from year to year. One of our research aims was to calculate future life expectancy at regional level. Life expectancy at birth is the average number of years that a newborn baby can expect to live, based on the prevailing mortality conditions in the year of their birth. The evolution of life expectancy, which can be calculated using current time series, indirectly reflects the impact of migration. People tend to migrate from regions with a lower life expectancy and less economic development to regions with a higher life expectancy and greater economic development. For example, higher earnings and changes in lifestyle can have a positive impact on migrants' life expectancy. The effects of various past events, including migration, are taken into account when predicting the mortality index(es) of the models used. Any intensive effects can be filtered out of the trend parameters if necessary.

The research topic is novel in the sense that multi-population models have not been applied to Hungarian regional data. However, multi-population models fitted to territorial data can be found in the international literature (see, e.g., Lee–Nault 1993, Danesi et al. 2015, Kleinow 2015, Enchev et al. 2017, Scognamiglio 2022). This research focuses on multi-population models with single and multiple mortality indices, which are among the best-known mortality models. A model with more mortality indices can consider periodic group-specific deviations from the main trend, as well as the long-term common trend. In this study, the fit and predictive accuracy of the five multi-population models were compared using various measures, and these models were also compared with the Lee–Carter model.

One of the multi-population models used is the Carter–Lee (1992) model, which includes a common mortality index and its age-varying group-specific coefficient. This does not necessarily give a coherent prediction. The $\beta_x^{(1)}\kappa_t^{(1)}$ factor in Li and Lee's (2005) CF model is common to subgroups. This is considered to be a coherent model where divergence does not become a problem in the long run. The ACF model (see Li–Lee 2005,

Li 2013) includes a subpopulation-specific mortality index in addition to the common factor. A coherent prediction can also be expected from this model due to the common factor. The CAE model proposed by Kleinow (2015) is a modification of the Lee–Carter model with a common age effect, but allowing for differences in the mortality index between subpopulations. The ACF–CAE model is obtained when the ACF model is modified for the coefficient of the additional time-varying factor on age so that it no longer differs between groups (see Li et al. 2016).

In this research, we used both adjusted and nonadjusted population data to see how much difference there was in the fitted parameter values of the models in the two cases. Using adjusted population data from the 1980s and 1990s, we found that the „outliers" were dampened in terms of mortality indices (see the online Appendix). This paper presents the results of models fitted to adjusted data. One of the eight Hungarian regions has been selected to illustrate the results. In terms of historical trends in life expectancy at birth, the Southern Great Plain is the region that best reflects the mortality of the country as a whole, for both men and women. The Southern Great Plain is considered a representative region. All figures and results for all other regions are available in the online Appendix. There were no convergence problems in fitting the models. However, the estimated parameter values of the Carter–Lee model were not accepted for Northern Hungarian women, so they were not included in the analysis and prediction.

The fitted parameters

Figures 6–7 show the fitted parameter values for the selected mortality models for men and women separately in the Southern Great Plain. In multi-population models, there is one or more time- and/or age-dependent parameter that is common to all Hungarian regions. This parameter represents the value for the country as a whole. Group-specific parameters reveal regional differences. The parameter α_{xj} is different for each model and region, and is only approximately equal to the mean of the base period log mortality rates because this parameter is optimized in each model fit, i.e., it is determined iteratively with the other parameters.[21] First, the mortality models were fitted to the national data by sex.

[21] The mean of the log mortality rates between 1970 and 2021 by age was chosen as initial values for the term α_{xj}.

In other words, they were treated as single-population models. The common parameter values were derived from these estimates, which were held fixed when fitting the multi-population variants.

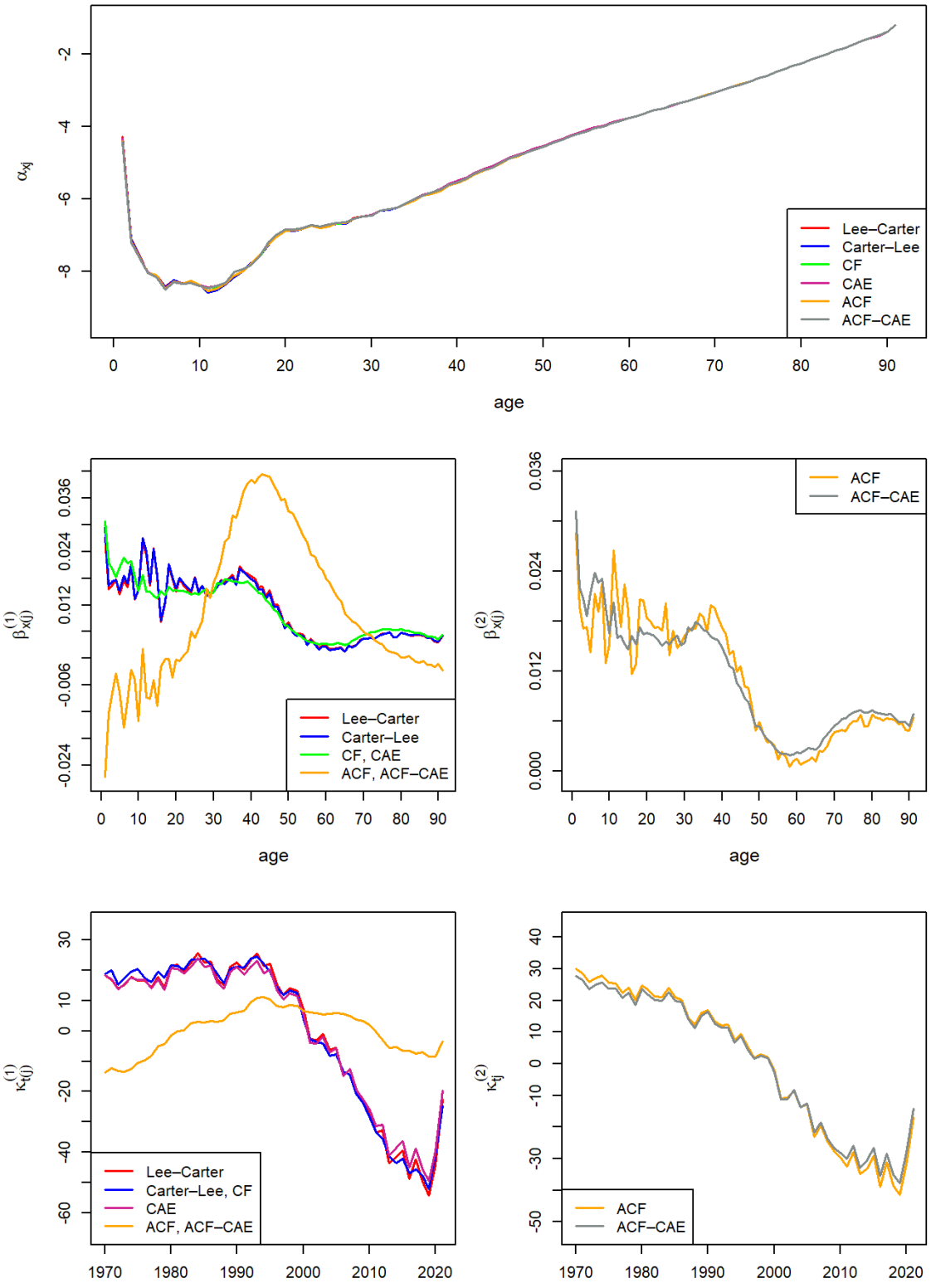
The long-term mortality trends described by the Lee–Carter, Carter–Lee, CF and CAE models show a high degree of similarity for men. Each of these models is a single-factor model. The multi-population variants do not differ significantly from the Lee–Carter model. Between 1970 and the mid-1990s, mortality did not improve and even slightly worsened for men, as shown by $\kappa_{t(j)}^{(1)}$. For men, the long-term improvement in mortality began after the change of political and economic regime in the second half of the 1990s. These four models show a different picture for women, for whom the long-term trend has been steadily improving since 1970, apart from small fluctuations from year to year. Based on the $\kappa_t^{(1)}$ parameter of the ACF and ACF–CAE models, not only men but also women show an initial decline in mortality over decades, followed by an improvement from the second half of the 1990s onwards. For these two models, the parameter $\kappa_{tj}^{(2)}$ indicates the deviation from the long-term trend, which has a negative slope overall, except in recent years. In 2020 and 2021, the COVID-19 epidemic caused a worsening in mortality for both men and women. For the ACF and ACF–CAE models, the second mortality indices appear to be predictable, and the paths of the first and second indices differ, confirming that it may be worth quantifying group-specific factors in addition to the national trend.

The values of the $\beta_{x(j)}^{(i)}$ parameters are not all positive across models and regions. Negative values appear at certain ages in the Lee–Carter model for women in Pest and for men in Northern Hungary. The Carter–Lee model also shows a negative value for 9-year-old females in Pest. In certain regions, the values of the $\beta_{xj}^{(2)}$ parameter in the ACF model are negative for both males and females. Negative values indicate increasing mortality in recent years, i.e., years in which the mortality index is already negative. If a model incorporates multiple factors, the estimated change in mortality rates across different age groups depends on all the mortality indices and their respective coefficients.

The parameter α_{xj} is the baseline shape of mortality. Like the $\beta_{x(j)}^{(i)}$ coefficient, it depends only on age. Women have lower α_{xj} values than men, indicating that women have better overall mortality rates than men. A remarkable difference can be observed in mortality between the sexes around the age of 20. The lower values for young men than

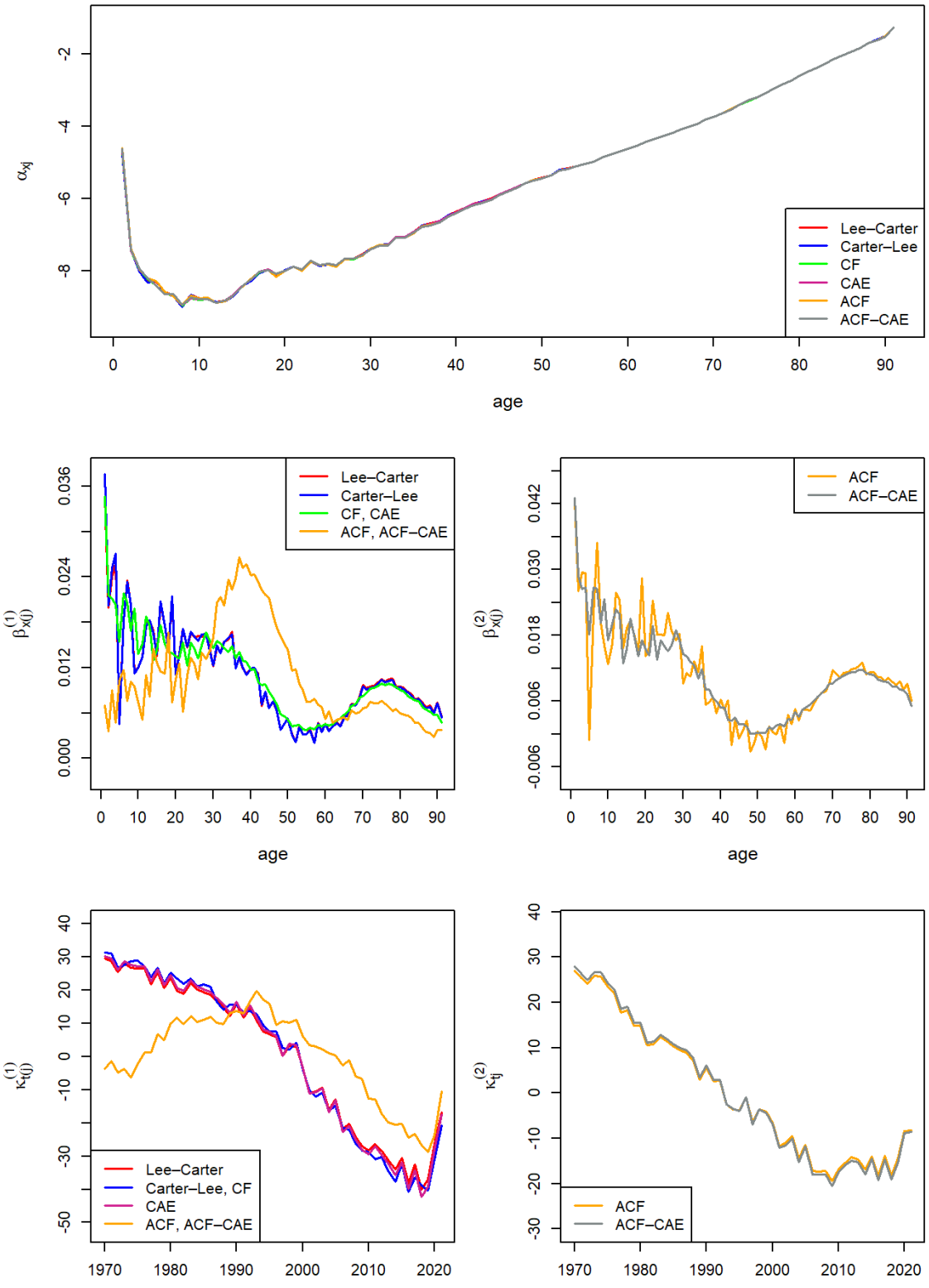
for women is mainly due to the newly obtained driving licence and more traffic accidents. In terms of parameter α_{xj}, there is almost no difference between the models.

Figure 6. The fitted parameters of the mortality models for males in the Southern Great Plain



Source: own calculation

Figure 7. The fitted parameters of the mortality models for females in the Southern Great Plain



Source: own calculation

The goodness of fit

The in-sample goodness-of-fit of the mortality models was assessed using AIC, BIC, MAPE, and the explanation ratio. Tables 3–4 show the values of these measures by model and region. Based on the in-sample analysis, the ACF model is the best choice for men, and the ACF–CAE model is the second best choice. For women, similar findings can be made on the basis of most metrics. We found that the additional factor decreased the AIC, BIC, and MAPE values, and increased the explanation ratio. Thus, the use of models with two mortality indices may be justified. The CF model for both sexes is one of the worst performing models.

Table 3. The mortality models' goodness of fit for males, 1970–2021

Regions	**Measures**	**Lee–Carter**	**Carter–Lee**	**CF**	**CAE**	**ACF**	**ACF–CAE**
Budapest	AIC	5,318,727	5,320,350	5,322,359	5,319,795	**5,313,249**	*5,314,690*
	BIC	5,320,226	5,321,849	5,323,859	5,321,294	**5,315,660**	*5,317,101*
	MAPE	0.225	0.227	0.237	0.228	**0.191**	*0.197*
	R	*0.499*	0.512	0.481	0.427	**0.580**	0.487
Central Transdanubia	AIC	3,172,025	3,172,114	3,172,185	3,172,103	**3,168,282**	*3,168,355*
	BIC	3,173,524	3,173,614	3,173,685	3,173,602	**3,170,692**	*3,170,765*
	MAPE	0.229	0.231	0.232	0.231	**0.193**	*0.195*
	R	0.382	0.376	0.372	0.378	**0.486**	*0.483*
Northern Great Plain	AIC	4,150,538	4,152,183	4,153,092	4,152,339	**4,146,637**	*4,147,339*
	BIC	4,152,037	4,153,682	4,154,592	4,153,839	**4,149,047**	*4,149,750*
	MAPE	0.240	0.230	0.236	0.237	**0.191**	*0.203*
	R	0.344	0.394	0.374	0.376	**0.521**	*0.492*
Northern Hungary	AIC	3,998,530	4,000,420	4,000,786	3,999,456	**3,992,801**	*3,993,114*
	BIC	4,000,030	4,001,920	4,002,285	4,000,955	**3,995,212**	*3,995,524*
	MAPE	0.251	0.238	0.241	0.237	**0.187**	*0.193*
	R	0.242	0.221	0.262	0.310	**0.477**	*0.464*
Pest	AIC	3,022,995	3,024,011	3,024,489	3,023,642	**3,019,225**	*3,019,943*
	BIC	3,024,495	3,025,510	3,025,988	3,025,141	**3,021,635**	*3,022,354*
	MAPE	0.240	0.254	0.256	0.246	**0.208**	*0.214*
	R	0.518	0.513	0.507	0.513	**0.587**	*0.582*
Southern Great Plain	AIC	4,384,221	4,384,292	4,384,632	4,384,487	**4,379,021**	*4,379,297*
	BIC	4,385,720	4,385,791	4,386,131	4,385,986	**4,381,431**	*4,381,708*
	MAPE	0.218	0.218	0.222	0.223	**0.180**	*0.186*
	R	0.443	0.440	0.427	0.433	**0.548**	*0.532*
Southern Transdanubia	AIC	3,068,996	3,069,075	3,069,282	3,069,208	**3,064,788**	*3,064,942*
	BIC	3,070,495	3,070,574	3,070,781	3,070,707	**3,067,199**	*3,067,352*
	MAPE	0.236	0.236	0.237	0.237	**0.198**	*0.199*
	R	0.350	0.344	0.340	0.335	**0.461**	*0.452*
	AIC	2,954,563	2,954,651	2,954,773	2,954,694	**2,951,070**	*2,951,213*

Western Transdanubia	BIC	2,956,062	2,956,150	2,956,272	2,956,193	**2,953,480**	*2,953,624*
	MAPE	0.243	0.244	0.245	0.244	**0.205**	*0.206*
	R	0.336	0.333	0.327	0.329	**0.441**	*0.433*

Note: The best values are highlighted in bold, and the second-best values are in italics.

Source: own calculation

Table 4. The mortality models' goodness of fit for females, 1970–2021

Regions	Measures	Lee–Carter	Carter–Lee	CF	CAE	ACF	ACF–CAE
Budapest	AIC	5,815,269	5,816,891	5,816,621	5,815,625	**5,814,146**	*5,814,635*
	BIC	*5,816,768*	5,818,390	5,818,120	5,817,124	**5,816,556**	5,817,045
	MAPE	0.222	0.238	0.242	0.231	**0.207**	*0.219*
	R	0.464	0.467	0.444	0.447	**0.504**	*0.479*
Central Transdanubia	AIC	2,706,458	2,706,541	2,706,661	2,706,583	**2,706,165**	*2,706,293*
	BIC	**2,707,957**	*2,708,040*	2,708,160	2,708,082	2,7085,75	2,708,703
	MAPE	0.236	0.236	0.236	0.237	**0.222**	*0.223*
	R	0.250	0.258	0.265	0.262	**0.304**	**0.304**
Northern Great Plain	AIC	3,903,348	3,903,488	3,903,683	3,903,426	**3,902,511**	*3,902,650*
	BIC	**3,904,847**	3,904,987	3,905,182	3,904,925	*3,904,921*	3,905,060
	MAPE	0.239	0.240	0.241	0.240	**0.221**	*0.224*
	R	0.339	0.330	0.340	0.349	*0.388*	**0.398**
Northern Hungary	AIC	3,523,379	–	3,524,586	3,523,570	**3,522,820**	*3,523,079*
	BIC	**3,524,879**	–	3,526,085	*3,525,070*	3,525,231	3,525,489
	MAPE	0.248	–	0.255	0.249	**0.230**	*0.234*
	R	0.195	–	0.127	0.193	**0.256**	*0.240*
Pest	AIC	2,640,869	2,641,056	2,641,247	2,641,158	**2,640,338**	*2,640,558*
	BIC	**2,642,369**	*2,642,555*	2,642,746	2,642,657	2,642,748	2,642,969
	MAPE	0.249	0.252	0.256	0.256	**0.236**	*0.240*
	R	0.343	0.345	0.325	0.327	**0.378**	*0.371*
Southern Great Plain	AIC	3,913,262	3,913,421	3,913,450	3,913,363	**3,912,915**	*3,913,010*
	BIC	**3,914,762**	3,914,920	3,914,949	*3,914,862*	3,915,325	3,915,421
	MAPE	0.235	0.236	0.237	0.237	**0.222**	*0.225*
	R	0.273	0.253	0.248	0.258	**0.314**	*0.296*
Southern Transdanubia	AIC	2,791,125	2,791,205	2,791,623	2,791,404	**2,790,788**	*2,790,989*
	BIC	**2,792,624**	*2,792,704*	2,793,122	2,792,903	2,793,198	2,793,399
	MAPE	0.240	0.241	0.241	0.240	**0.227**	*0.229*
	R	0.220	0.217	0.222	0.193	**0.259**	*0.232*
Western Transdanubia	AIC	*2,630,525*	2,630,669	2,630,891	2,630,653	**2,630,405**	2,630,550
	BIC	**2,632,024**	2,632,169	2,632,390	*2,632,152*	2,632,816	2,632,960
	MAPE	0.234	0.235	0.236	0.235	**0.222**	*0.224*
	R	0.259	0.276	0.262	0.232	**0.299**	*0.267*

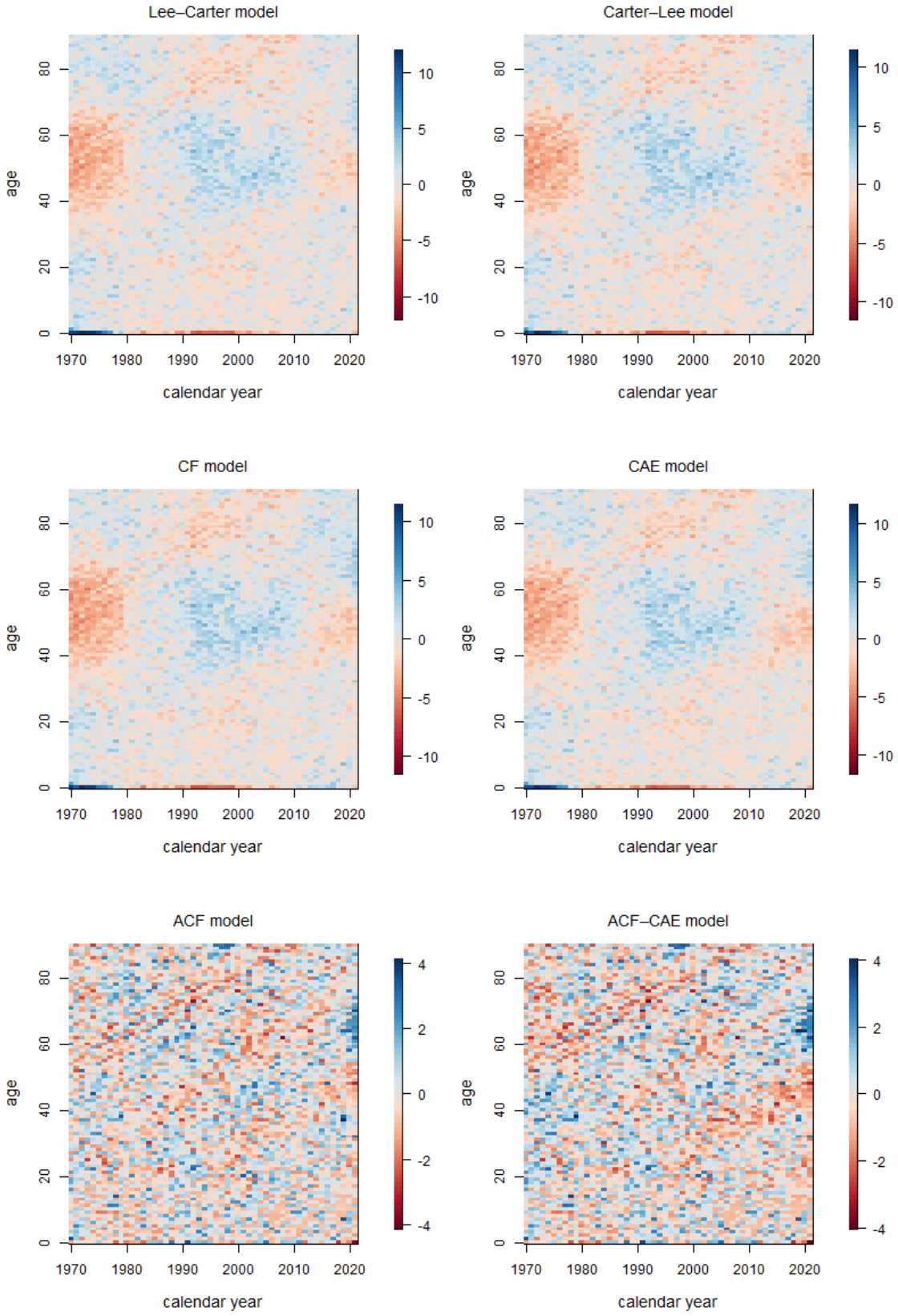
Note: The best values are highlighted in bold, and the second-best values are in italics.

Source: own calculation

The standardized residuals were also examined after fitting the mortality models. These values were presented using heatmaps and scatter plots.[22] Heatmaps show the structure of the residuals by calendar year on the x-axis and by age on the y-axis. The better the model, the more random the arrangement and the smaller the values of the standardized residuals. The patterns in the figures suggest that the models have not been able to quantify all the effects. For example, the diagonal pattern highlights the need to account for the cohort effect. No models with cohort effects were fitted in the regional analysis of Hungarian mortality. One solution to the systematic patterns in the standardized residuals may be to include additional time-dependent parameters. Cohort effects and additional mortality indices may improve the model fit. The ACF and ACF–CAE models provide a better fit than the other models, as illustrated by the heatmaps (see Figures 8–9). The scatter plots show the residuals as a function of age, calendar year and year of birth (see Figures 10–11), confirming that model extension improves the estimation. The online Appendix contains further figures for other regions.

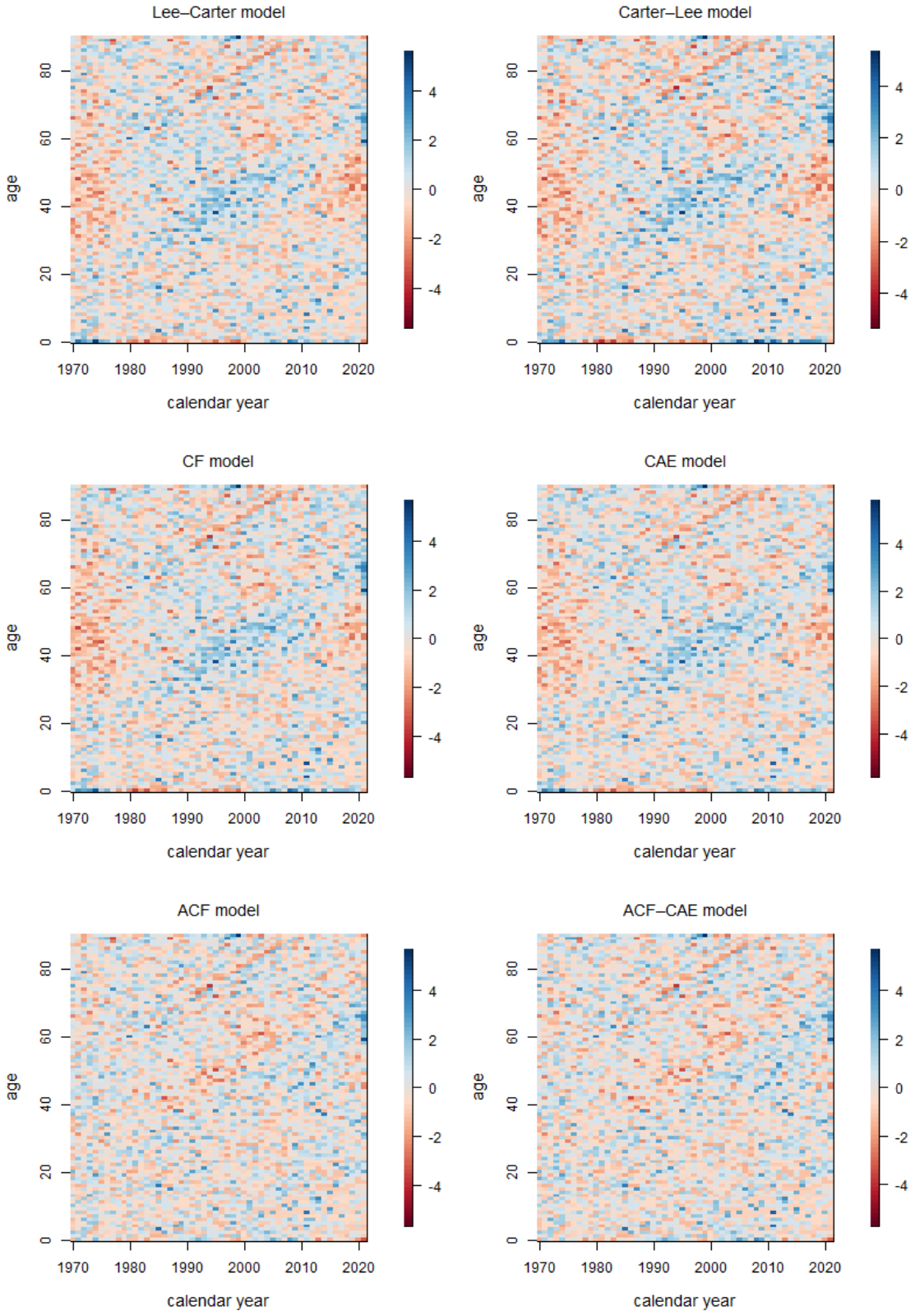
[22] We applied the *StMoMo* package of Villegas et al. (2018) to prepare these figures.

Figure 8. The standardized residuals of the mortality models for males in the Southern Great Plain



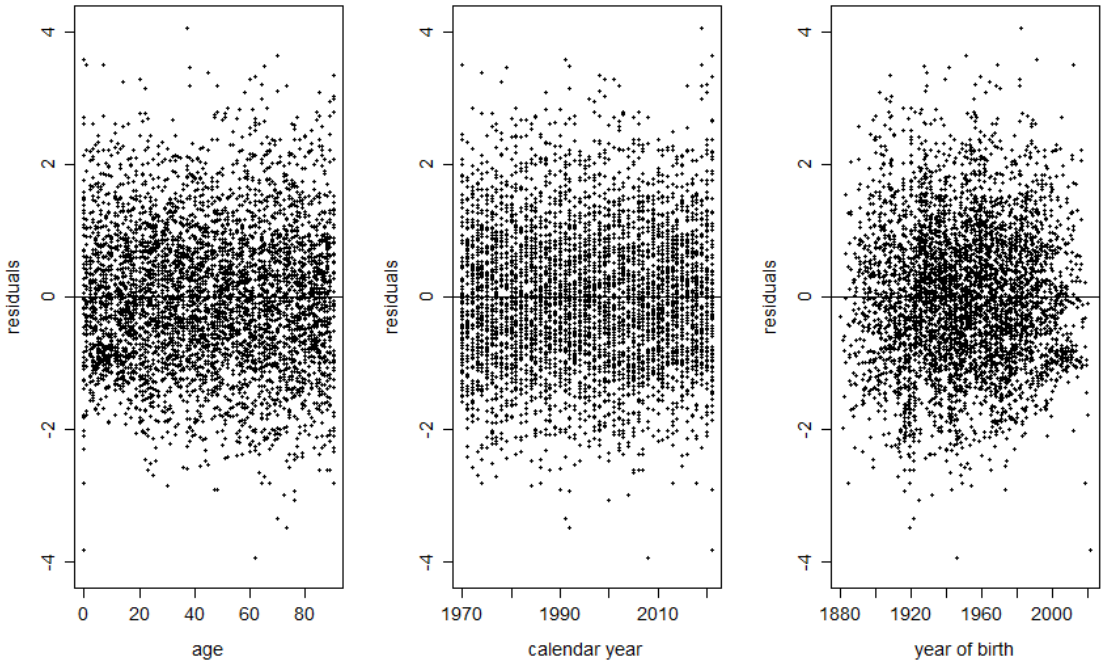
Source: own calculation

Figure 9. The standardized residuals of the mortality models for females in the Southern Great Plain



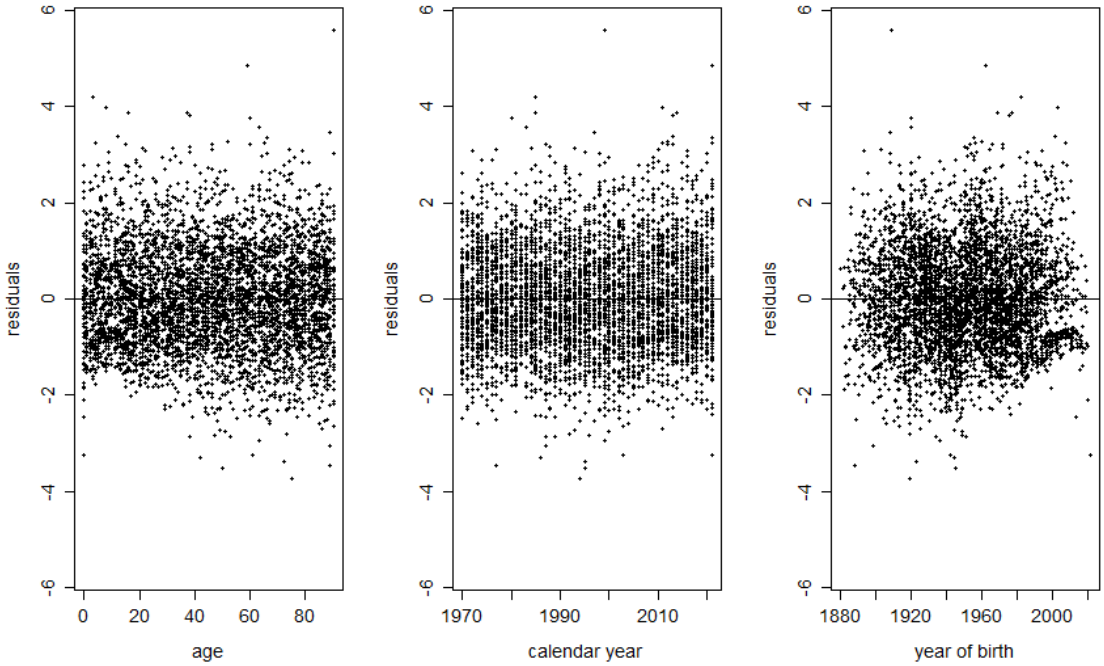
Source: own calculation

Figure 10. The standardized residuals of the ACF model for males in the Southern Great Plain



Source: own calculation

Figure 11. The standardized residuals of the ACF model for females in the Southern Great Plain



Source: own calculation

Forecasting the mortality indices

To forecast future mortality and calculate life expectancy, it is necessary to forecast the mortality index(es). Projections up to 2050 were made using ARIMA models. We tested for the presence of a unit root (with the augmented Dickey–Fuller and Kwiatkowski–Phillips–Schmidt–Shin tests), normality (with the Shapiro–Wilk test), and autocorrelation of the residuals (with the Ljung–Box Q-statistic and the Breusch–Godfrey test) during the time series analysis.[23] The values of the information criteria (AIC and BIC)[24] for the possible ARIMA models were examined. Based on these and the above tests, we selected the appropriate ARIMA models for predicting mortality indices. The results of the test statistics and the AIC and BIC values can be found in the online Appendix.

In forecasting the mortality indices, we have assumed their independence between regions on the one hand, and their independence from each other in the case of two-factor models on the other hand. Tables 5–6 summarize which ARIMA models were found to be the best fit for each mortality index. For each region, the same ARIMA model(s) could be used for most of the mortality models. The exception to this is the Lee–Carter model. For males, different models were used to forecast mortality in two regions: the Northern Great Plain and Northern Hungary. Figures 12–13 show the predicted values of the mortality indices and their 80% and 95% prediction intervals for the ACF model by sex. Mortality was predicted to improve until 2050.

The impact of the COVID-19 epidemic in 2020 and 2021 was filtered out using exogenous dummy variables. The projection was made under two scenarios: 1) mortality is expected to be higher in 2022 (dummy variable set to one), 2) no excess deaths are expected in the first year of the projection and thereafter (dummy variable set to zero). The predicted mortality indices for the other models and the results under the two scenarios can be found in the online Appendix. In the following, we present the results for the second case. In addition to the predicted values of the mortality indices, the fitted age-varying parameter values (α_{xj} and $\beta_{x(j)}^{(i)}$) were required to calculate future mortality rates. The error terms were set to zero in the forecast.

---

[23] The *stats* (R Core Team 2021), the *tseries*, and the *lmtest* packages were used for the testing (see https://CRAN.R-project.org/package=tseries and https://CRAN.R-project.org/package=lmtest).

[24] The formulas are as follows: $AIC = T\ ln(sum\ of\ squared\ residuals) + 2n$ and $BIC = T\ ln(sum\ of\ squared\ residuals) + n\ ln(T)$ where n is the number of estimated parameters ($p + q +$ possible constant term), and T is the number of usable observations (see, e.g., Enders 2014).

An important issue in the forecast of mortality indices is the treatment of structural breaks when they occur in the time series. Coelho and Nunes (2011) introduced structural break tests, following Harvey et al. (2009) and Harris et al. (2009), and predicted the mortality index of the Lee–Carter model in two ways: with and without allowing for a structural break. Coelho and Nunes (2011) tested for structural breaks, while Sobreira and Nunes (2016) used a test for multiple breaks. In their analysis, Coelho and Nunes (2011) found that structural breaks were more frequent for men than for women. If mortality improvement accelerates after the structural break, then higher life expectancy can be predicted when the structural break is taken into account. The methods of Coelho–Nunes (2011) and Sobreira–Nunes (2016) were not applied in this study, but their approaches represent possible directions for further research. Li et al. (2011) presented the procedure of Zivot–Andrews (1992) to identify structural breaks in the mortality index of the Lee–Carter model. This procedure is suitable for distinguishing between a random walk with drift and a broken-trend stationarity process. Since the ARIMA model chosen for forecasting is often different from the random walk with drift process, this test was not used in our research.

Table 5. ARIMA models of the mortality indices for males

The mortality models	$\kappa_{t(j)}^{(1)}$	$\kappa_{tj}^{(2)}$	**Regions**
Lee–Carter	ARIMA(0,1,0) with drift, ARIMA(0,1,0), ARIMA(0,1,1) with drift	–	Northern Great Plain, Northern Hungary, other regions
Carter–Lee	ARIMA(0,1,1) with drift	–	all regions
CF	ARIMA(0,1,1) with drift	–	all regions
CAE	ARIMA(0,1,1) with drift	–	all regions
ACF	ARIMA(0,1,0)	ARIMA(0,1,1) with drift	all regions
ACF–CAE	ARIMA(0,1,0)	ARIMA(0,1,1) with drift	all regions

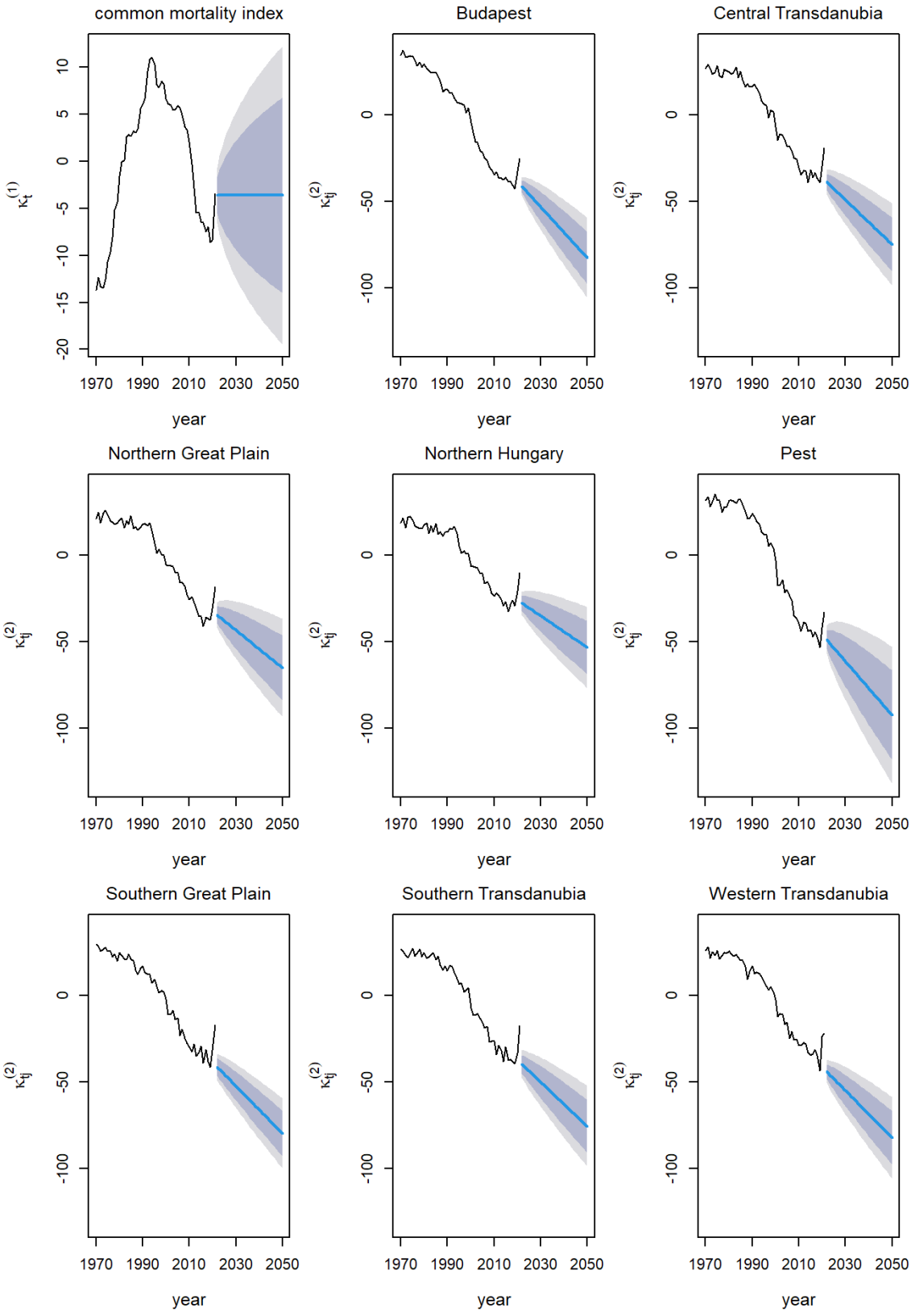
Source: own editing

Table 6. ARIMA models of the mortality indices for females

The mortality models	$\kappa_{t(j)}^{(1)}$	$\kappa_{tj}^{(2)}$	**Regions**
Lee–Carter	ARIMA(0,1,1) with drift	–	all regions
Carter–Lee	ARIMA(0,1,1) with drift	–	all regions
CF	ARIMA(0,1,1) with drift	–	all regions
CAE	ARIMA(0,1,1) with drift	–	all regions
ACF	ARIMA(0,1,0) with drift	ARIMA(0,1,1) with drift	all regions
ACF–CAE	ARIMA(0,1,0) with drift	ARIMA(0,1,1) with drift	all regions

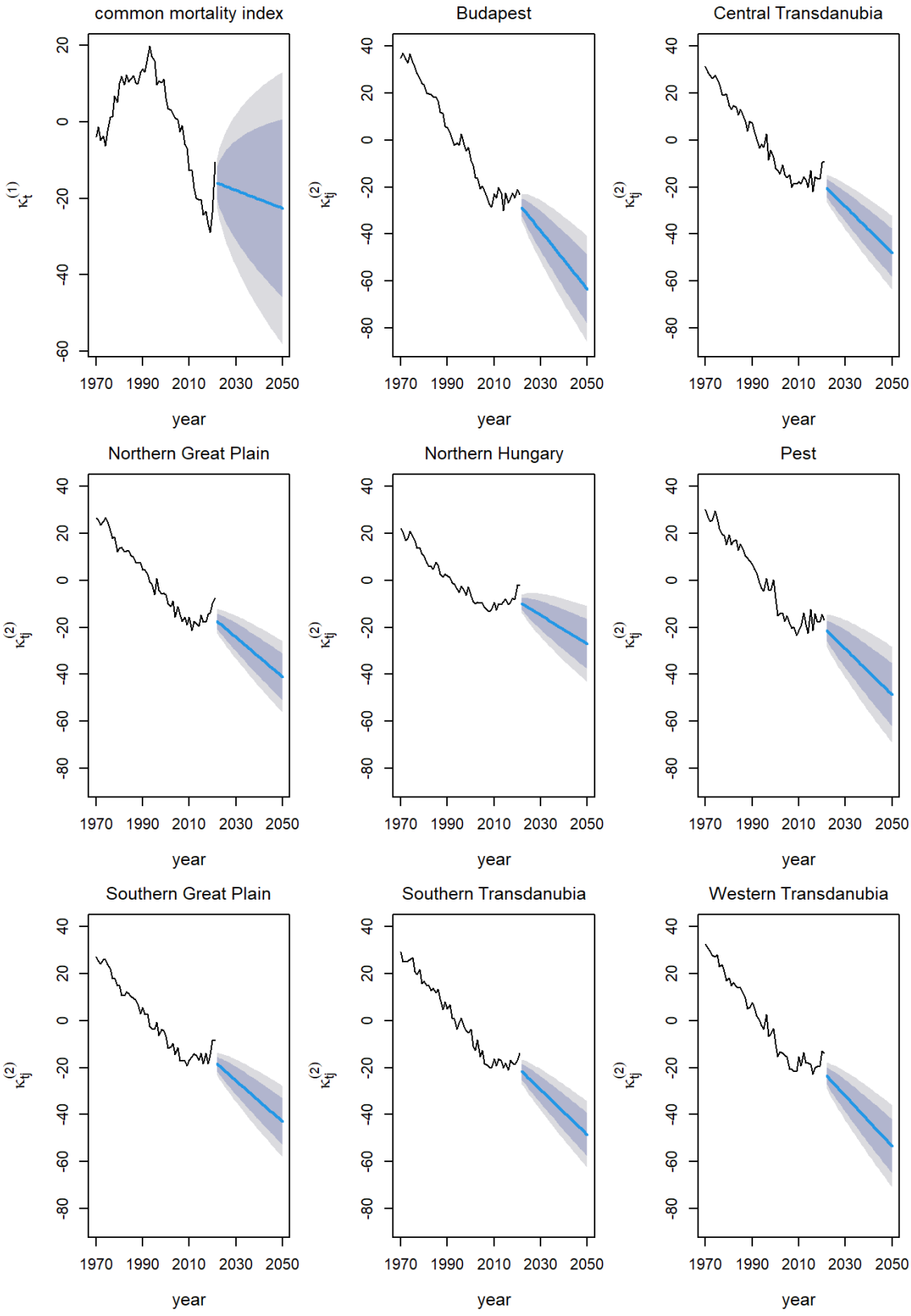
Source: own editing

Figure 12. The common and group-specific mortality indices of the ACF model for males



Note: The gray backgrounds represent the 80% and 95% prediction intervals.
Source: own calculation

Figure 13. The common and group-specific mortality indices of the ACF model for females



Note: The gray backgrounds represent the 80% and 95% prediction intervals.
Source: own calculation

The future life expectancy at birth by region

The projected mortality rates can be used to calculate the life table. Figures 14–15 show the evolution of life expectancy at birth[25] over the next nearly 30 years according to different mortality models. It is assumed that life expectancy will return to the level before the COVID-19 pandemic in the near future and then continue to improve in the long term. We can examine how well each model meets this expectation on the basis of the projected values, i.e., when life expectancy will reach the level of 2019.

In the Carter–Lee and CF models, the long-term trend in mortality change is the same for all regions. Thus, there are smaller regional differences in future life expectancy at birth than in the other models. The Carter–Lee and CF models estimate the lowest life expectancy in 2050 for men. The results of the Lee–Carter and CAE models are more favourable. The ACF and ACF–CAE models give the most optimistic predictions for men. According to these mortality models, longevity increases measurably by 2050 in Budapest, Central Transdanubia, the Northern Great Plain, Pest, the Southern Great Plain, and Southern Transdanubia. Based on the two-factor models, the regional projections of male life expectancy at birth in 2050 are as follows: 78.3–78.9, 74.6–74.6, 74.7–74.7, 75.4–75.9, 74.7–74.8, and 74.8–74.8 years. In Northern Hungary, which is characterized by unfavourable mortality conditions, the ACF–CAE and CF models predict 72.1–72.3 years for men in 2050. These are the most optimistic estimates for this region. Life expectancy at birth at the end of the time horizon will be 76.4–76.5 years for men in Western Transdanubia according to the Lee–Carter and CAE models. The other models estimate a lower life expectancy.

For women, the picture is different: the ACF and ACF–CAE models do not provide the most favourable predictions. The CAE model is considered the most optimistic, while the two-factor models are considered the most pessimistic. According to the CAE and Lee–Carter mortality models, female life expectancy at birth in 2050 will be 84.0–84.2, 82.8–82.8, 82.5–82.7, 82.7–82.9, and 83.8–83.9 years in Budapest, Central Transdanubia, Pest, Southern Transdanubia, and Western Transdanubia, respectively. In the Northern Great Plain and Northern Hungary, life expectancy will increase to 82.0–82.4 years and 81.0–82.4 years, respectively by 2050, based on the optimistic predictions

[25] In this study, life tables were calculated using the methodology of Chiang (1968) and Eurostat (see https://ec.europa.eu/eurostat/cache/metadata/Annexes/demo_mor_esms_an_1.pdf), except that the group of people aged 90 and over was considered as the oldest cohort instead of people aged 85 and over.

of the Carter–Lee and CF models. According to the Carter–Lee and CAE models, 82.6–82.7 years is the most favourable projection for women in the Southern Great Plain in 2050.

Three of the six models show that male life expectancy will reach the level of 2019 in all regions. The optimistic ACF and ACF–CAE models predict this will happen in 2035–2036, while the CAE model estimates 2048. The other three models predict that in at least two regions life expectancy will not even reach 2019 levels by 2050. The conclusion for women is different, with all models projecting that future life expectancy will exceed the 2019 level within the next few years in all regions. According to the Lee–Carter and CAE models, this will occur in 2025 or 2027. The ACF–CAE model predicts that life expectancy for women in all regions will exceed pre-pandemic levels by 2033.

Using multi-population models, we expected that there would be no significant change in the ranking of regions by life expectancy at birth over the forecast period. Based on the average ranking between 2010 and 2021, life expectancy was highest for both men and women in Budapest and lowest in Northern Hungary. For men, the Northern Great Plain, Pest, and Western Transdanubia followed Budapest in the same rank, while Central Transdanubia, the Southern Great Plain and Southern Transdanubia had higher life expectancy than Northern Hungary. The second highest life expectancy for women between 2010 and 2021 was observed in Western Transdanubia. This region was followed by Central Transdanubia, Pest, and the Southern Great Plain. In the Northern Great Plain and Southern Transdanubia, life expectancy for women has been higher in recent years than in Northern Hungary.

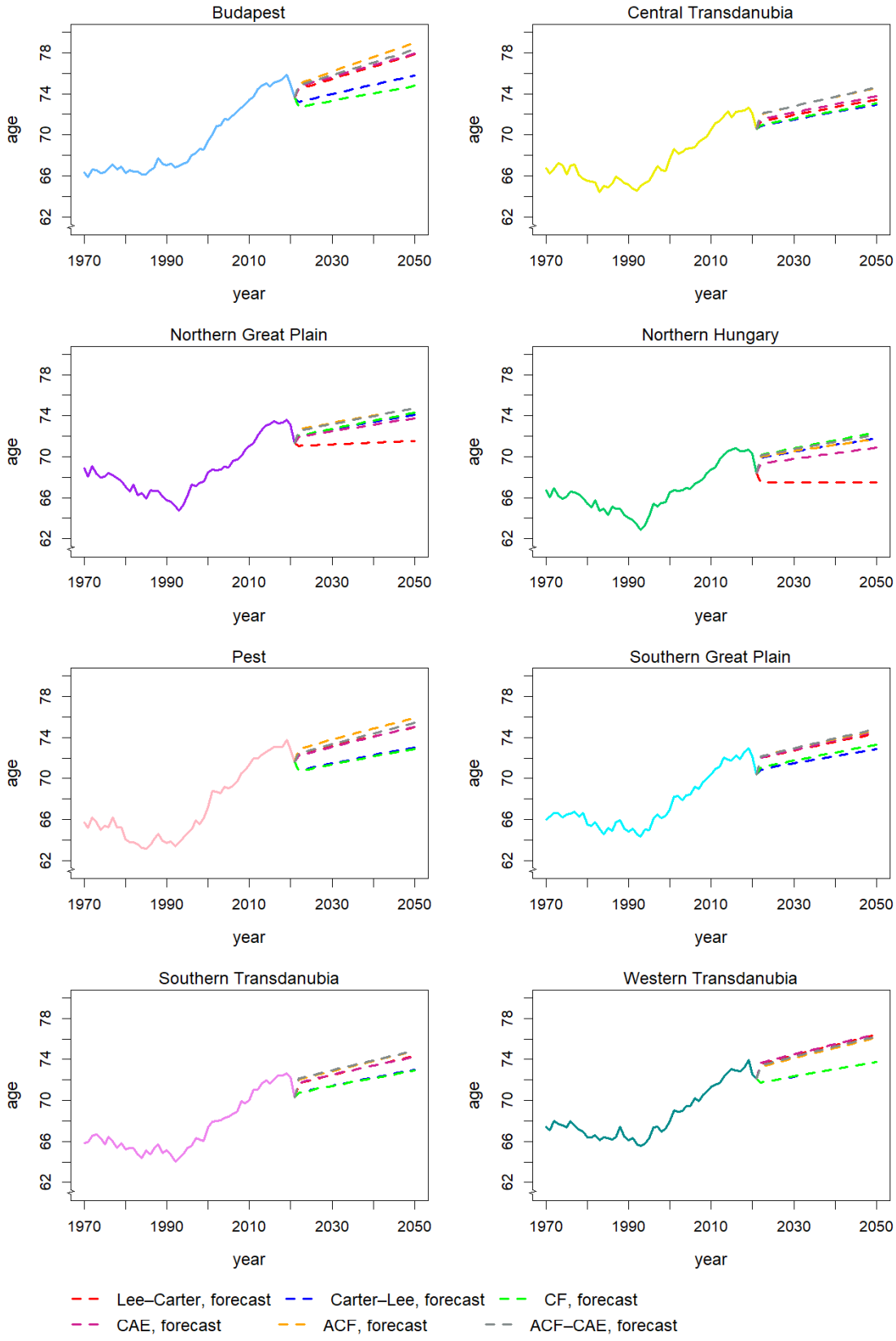
The Northern Great Plain is ranked low in the men's future rankings, coming 7th in the rankings based on the Lee–Carter model predictions. According to the Carter–Lee model, it is not realistic for Pest to be ranked 6th for the next 10 years. The CF model reaches a similar conclusion: Pest will rank 7th until 2049. With regard to the CAE model, the position of the Northern Great Plain is questionable. In the case of the ACF and ACF–CAE models, the decline of the Northern Great Plain, and the improvement of Southern Transdanubia in the last years of the forecast do not correspond to the trend observed in the past.

Analyzing the results for women, the ranking predicted by the Lee–Carter model does not show any significant reordering across regions. In the long term, the Carter–Lee model estimates a lower ranking for Pest, but an improved ranking for Southern Transdanubia and the Southern Great Plain. A switch between Budapest and Western

Transdanubia does not seem realistic. According to the projections of the CF model, Southern Transdanubia is the worst performing region, not Northern Hungary. Based on the CAE model, Pest is ranked lower, while life expectancy at birth in Southern Transdanubia will be higher in the long term than in the period 2010–2021. Southern Transdanubia is ranked 3rd on the basis of the CAE model. The conclusions of the ACF and ACF–CAE models are similar to those of the Carter–Lee and CAE models. Pest falls to 6th place, while Southern Transdanubia moves up to 4th place. However, this seems less realistic. Data on current and future life expectancy at birth, as well as the regions' rankings, are also available in the online Appendix.

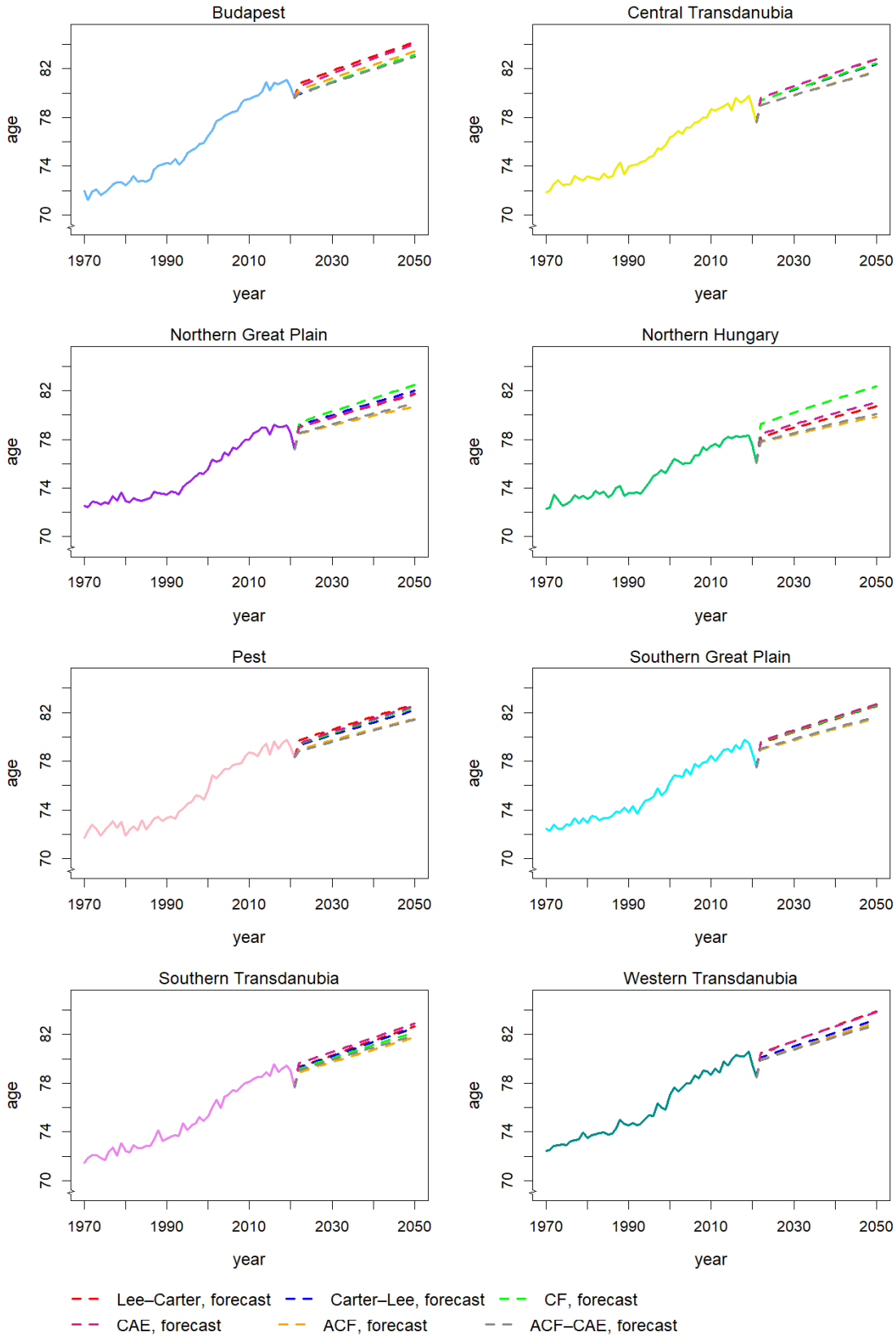
It is interesting to compare men and women in terms of life expectancy at birth. In 2019, Budapest and the Northern Great Plain had the smallest difference in life expectancy between men and women: 5.2 and 5.5 years, respectively. The largest difference was observed in Central Transdanubia and Northern Hungary: 7.1 and 7.6 years, respectively. The ACF and ACF–CAE models show a smaller or almost identical difference in the long-term forecast than was observed in 2019. The difference in life expectancy between men and women will continue to decrease in the future in Budapest according to the ACF and ACF–CAE models, in Pest according to the ACF model, and in Western Transdanubia according to the ACF–CAE model. The other models predict greater differences between the sexes in each region, i.e., life expectancy for women will increase relative to men.

Figure 14. Regional life expectancy at birth for males between 1970 and 2050



Source: own calculation

Figure 15. Regional life expectancy at birth for females between 1970 and 2050



Source: own calculation

Out-of-sample analysis

Based on the in-sample analysis, we previously found that the ACF and ACF–CAE models had the best fit. The same finding can be made on the basis of the out-of-sample analysis where a shorter base period was selected (see Tables 7–8). We fitted all mortality models for the period 1970–2015 and projected mortality rates for the period 2016–2019, so that estimated and actual mortality rates could be compared. The MAPE measure was applied for comparison. For the out-of-sample analysis, the same ARIMA models were used as in Tables 5 and 6. The ACF and ACF–CAE models were found to be the best performers for both men and women in almost all regions. For men in Pest and women in Budapest, the ACF and Lee–Carter models were more accurate than the ACF–CAE model. More results are available in the online Appendix.

Table 7. MAPE values of the mortality models for males, 2016–2019

Regions	Lee–Carter	Carter–Lee	CF	CAE	ACF	ACF–CAE
Budapest	0.358	0.330	0.350	0.299	*0.275*	**0.225**
Central Transdanubia	0.264	0.273	0.262	0.264	*0.217*	**0.209**
Northern Great Plain	0.515	0.284	0.306	0.315	**0.216**	*0.243*
Northern Hungary	0.580	0.283	0.249	0.255	**0.218**	*0.219*
Pest	*0.246*	0.341	0.384	0.350	**0.227**	0.262
Southern Great Plain	0.274	0.275	0.289	0.308	**0.224**	*0.245*
Southern Transdanubia	0.301	0.292	0.272	0.269	*0.214*	**0.203**
Western Transdanubia	0.313	0.309	0.286	0.275	*0.232*	**0.215**

Note: The best values are highlighted in bold, and the second-best values are in italics.
Source: own calculation

Table 8. MAPE values of the mortality models for females, 2016–2019

Regions	Lee–Carter	Carter–Lee	CF	CAE	ACF	ACF–CAE
Budapest	*0.297*	0.372	0.402	0.355	**0.253**	0.303
Central Transdanubia	0.294	0.295	0.288	0.288	*0.254*	**0.247**
Northern Great Plain	0.278	0.287	0.264	0.268	*0.239*	**0.225**
Northern Hungary	0.268	0.273	0.239	0.235	*0.216*	**0.204**
Pest	0.289	0.300	0.328	0.333	**0.250**	*0.270*
Southern Great Plain	0.259	0.254	0.238	0.239	*0.202*	**0.196**
Southern Transdanubia	0.285	0.286	0.254	0.246	*0.233*	**0.225**
Western Transdanubia	0.284	0.288	0.277	0.266	*0.229*	**0.225**

Note: The best values are highlighted in bold, and the second-best values are in italics.
Source: own calculation

Coherence analysis

The projections from the multi-population models can be evaluated in terms of whether they produce a coherent projection across regions. To test this, we compared age-specific regional mortality rates by sex, i.e., we calculated region-to-region ratios. The regional forecast is coherent if these ratios remain constant and do not diverge. Li (2013) and Li et al. (2016) also used such a test in their research.

This study shows that the CF model is clearly coherent. That means it supports the findings of Li–Lee (2005). The results of the calculations related to the coherence analysis can be found in the online Appendix. According to the region-to-region ratios, the ACF–CAE and CAE models do not show coherence in all age groups. For men, coherence is reached from the age of 45, while for women it is reached from the age of 40. Wen et al. (2021) found that the mortality models with common age effect(s) provided a coherent prediction. According to Scognamiglio (2022), the ACF model is a coherent model. This was not confirmed for all ages in this research. The Lee–Carter and the Carter–Lee models had divergence problems between regional mortality rates in this analysis. Enchev et al. (2017) found that nondivergent forecasting could be achieved using a multivariate time series model for the mortality models under investigation.

Summary

In this study, five multi-population mortality models and the Lee–Carter model were tested on Hungarian regional data, and we projected life expectancy at birth by sex. The models were evaluated according to several criteria. In-sample analysis showed that the ACF and ACF–CAE were the best fitting models for both males and females. These models also performed well for men in terms of regional ranking based on future life expectancy. In addition, the augmented models reduced or maintained the gap in life expectancy between men and women by region. This was due to the ACF and ACF–CAE models being the most optimistic for men and the most pessimistic for women. Overall, the augmented models performed best for men. For women, the Lee–Carter model was the most appropriate in terms of future life expectancy and long-term regional ranking.

One of the difficulties of forecasting was the significant impact of the COVID-19 epidemic on mortality over the past two years of the current time series. Therefore, forecasting for the first year of the forecast period was also challenging. According to the

literature, life expectancy at birth is expected to improve more in the years following the pandemic because of the excess mortality that primarily affected older age groups (see, e.g., Cairns et al. 2020, Goldstein–Lee 2024). For mortality models that predict lower life expectancies than expected, one solution is to adjust the baseline shape of mortality when making projections, while also removing the effect of the epidemic. The idea of modifying the value of α_{xj} was originally proposed by Lee–Miller (2001). This enables a level shift in life expectancy at birth, allowing us to use a parameter value that predicts a more realistic life expectancy in the early years after the COVID-19 pandemic. The main objective of the research was not to accurately assess the impact of the pandemic, but rather to demonstrate how multi-population models behave when applied to Hungarian regional data.

Moreover, the second half of the 1990s saw the start of a mortality improvement for men in particular, which was intended to offset the decline in mortality of previous decades. If we only consider the trend in the years following the structural break, favourable life expectancy values may be obtained for the forecast years that are not necessarily plausible, given that the rate of improvement in mortality rates from previous unfavourable levels may slow down in the long term.

In Hungary, the differences in mortality rates between regions justify the use of multi-population models with multiple mortality indices. This allows short-term differences to be considered alongside the common long-term trend. Models with a single, common mortality index reduce regional differences in forecasting, so their use should be avoided if this is not the expectation in the long run. However, these models, such as the CF model, provide greater coherence (see Li–Lee 2005).

If multi-population mortality models with a common mortality index are not good predictors, they should not necessarily be excluded. They may represent a possible prediction scenario if a subpopulation does not follow the pattern of the total population in the base period, but is assumed to do so in the future. A possible research direction is to consider the cohort effect in each model, which could improve the fit of the models. Yang et al. (2016) extended multi-population models with cohort effects: they considered both group-specific and common variants. This extension can also increase the predicted life expectancy at birth, providing an additional opportunity for more accurate predictions following a shock, such as the impact of the COVID-19 pandemic.

In our research, we fitted some basic multi-population mortality models that have not yet been presented using Hungarian data by other researchers. When using multi-

population models to predict mortality, we disregarded the following extensions: predicting the mortality index while accounting for a structural break (see Coelho–Nunes 2011, Sobreira–Nunes 2016), modifying the baseline shape of mortality (see Lee–Miller 2001), considering the cohort effect (see Renshaw–Haberman 2006, Yang et al. 2016), and using a multivariate time series model (see Enchev et al. 2017). However, the projected life expectancy at birth values suggest that further extensions should be considered to improve the models' predictive accuracy.

This research could also be improved in terms of how subpopulations are selected. Based on the explanation ratio, Li–Lee (2005) suggested testing whether it is necessary to exclude a member of a subgroup. If it is appropriate to restrict or possibly expand a subpopulation, the model parameters should be re-estimated. Following the idea of Kleinow (2015), parameters $\beta_{xj}^{(i)}$ could be grouped by age group or subpopulation based on cluster analysis, so that we could use common parameter values for a given age group or subpopulation. Similarly, clustering of $\kappa_{tj}^{(i)}$ parameters may be a feasible idea: first fit single-population mortality models, then cluster their time-varying parameters to form subpopulations, and fit multi-population mortality models. The latter was tested in the doctoral research: we clustered the mortality indices by cause of death. The results are described in the next chapter.

IV.2. Forecasting cause-of-death mortality with single- and multi-population models in Hungary[26]

Introduction

Cause-specific mortality in Hungary was also examined by selecting models from the Lee–Carter model family. We selected single- and multi-population models that made different assumptions about the distribution of deaths. These models also differed in the number of parameters in their initial equations, i.e., the number of mortality indices and age-dependent coefficients. One novel feature of the analysis is that multi-population models were developed by clustering time-varying indices from single-population models. The dynamic time warping algorithm was applied to measure the distance between mortality indices for the purpose of clustering. Different clusters of causes of

[26] Varga (2025) is the basis for this chapter.

death were formed for the two sexes, probably due to different lifestyles and mortality trends. Mortality models were fitted using data on people aged 60 and over from 1970 to 2021, disaggregated by sex and primary cause of death.[27] Several models were used to project cause-specific ASMR values to 2050.

First, the examined single-population models from the Lee–Carter family were fitted separately for each cause of death and sex. The following seven mortality models were selected: the Lee–Carter model and its two-factor version with both Poisson and Binomial distributions (see Booth et al. 2002, Brouhns et al. 2002, Renshaw–Haberman 2003, Villegas et al. 2018), the Binomial CBD model (see Cairns et al. 2006, Villegas et al. 2018), and two variants of the reduced Plat model assuming a Poisson distribution (see Plat 2009). Our goals when selecting single-population models were twofold: first, to select models that had not yet been examined in previous research using Hungarian data, and second, to become more familiar with models featuring predefined age-dependent parameters.

Multi-population models were used to examine the similarities in mortality rates across different subgroups (causes of death). Multi-population models have common time- or age-dependent parameters. These are the same for all subpopulations (see, e.g., Carter–Lee 1992, Li–Lee 2005, Kleinow 2015, Li 2013, Li et al. 2016, Enchev et al. 2017, Wen et al. 2021). Multi-population models are frequently fitted to data disaggregated by sex or territory (see, e.g., Lee–Nault 1993, Danesi et al. 2015, Kleinow 2015, Enchev et al. 2017, Scognamiglio 2022, Varga 2023). The common parameters of the multi-population models were determined by clustering the time-dependent mortality indices of the single-population models. Predictions were made using both single-population and multi-population models, based on mortality data by primary cause of death. The different models were compared based on their in-sample goodness of fit and out-of-sample predictive accuracy, with the aim of determining the most suitable model for use. The results of the research include identifying which primary causes of death are predicted to worsen or improve in the future, and whether the mortality gap between the sexes will increase or decrease.

[27] For this analysis, cause-specific deaths were divided by total population, separately for each sex. This means that the initial and central exposures were not specific to any particular cause of death. It is impossible to know how many people in the population are at risk of dying from a certain cause. Furthermore, individuals are exposed to multiple risks simultaneously. If we calculate the ratio of cause-specific deaths to total population for each cause, and then add these ratios together, we get the total death ratio to total population.

Using multi-population models to analyze mortality data by cause of death is a novel approach. Aggregate rates are often considered, and methods such as vector autoregressive models, vector error correction models and copulas are used to capture the relationship between causes of death and make forecasts (see, e.g., Arnold–Sherris 2013, 2015, Arnold–Glushko 2021, Carriere 1994, Dimitrova et al. 2013, Kaishev et al. 2007, Li–Lu 2018, Zittersteyn–Alonso-García 2021). The impact of changes in lifestyle, advances in the medical field or epidemics can be even more striking when mortality is analyzed by main cause of death, with the results showing how these factors can influence mortality rates. The reason for this is that these effects do not necessarily affect mortality in the same way at the same time. If mortality trends for different causes of death are similar, these causes may be clustered together.

Notable parametric and nonparametric mortality models[28] were chosen for analyzing and predicting cause-specific mortality. We selected the CBD and reduced Plat models from among the models with predefined age-dependent variables. We also fitted multi-population versions of these single-population models, developed by clustering the mortality indices. Of these multi-population models, only the Carter–Lee (1992) model is referenced in the literature. An innovation of this research is the creation of multi-population versions of the parametric CBD and reduced Plat single-population models. Furthermore, the present study contributes to the literature by creating multi-population variants based on a cluster analysis of mortality indices, employing the DTW distance measure. Using the DTW algorithm is a new approach in this research area.

For the analysis, we selected the main groups of causes of death that accounted for the majority of deaths. The nine main groups selected are listed in order of ICD classification as follows:

I Certain infectious and parasitic diseases,

II Neoplasms,

IV Endocrine, nutritional and metabolic diseases,

VI Diseases of the nervous system,

IX Diseases of the circulatory system,

X Diseases of the respiratory system,

XI Diseases of the digestive system,

XIV Diseases of the genitourinary system, and

[28] In nonparametric models, all parameters have to be fitted. In parametric models, there are predefined age-dependent variables in the initial equations.

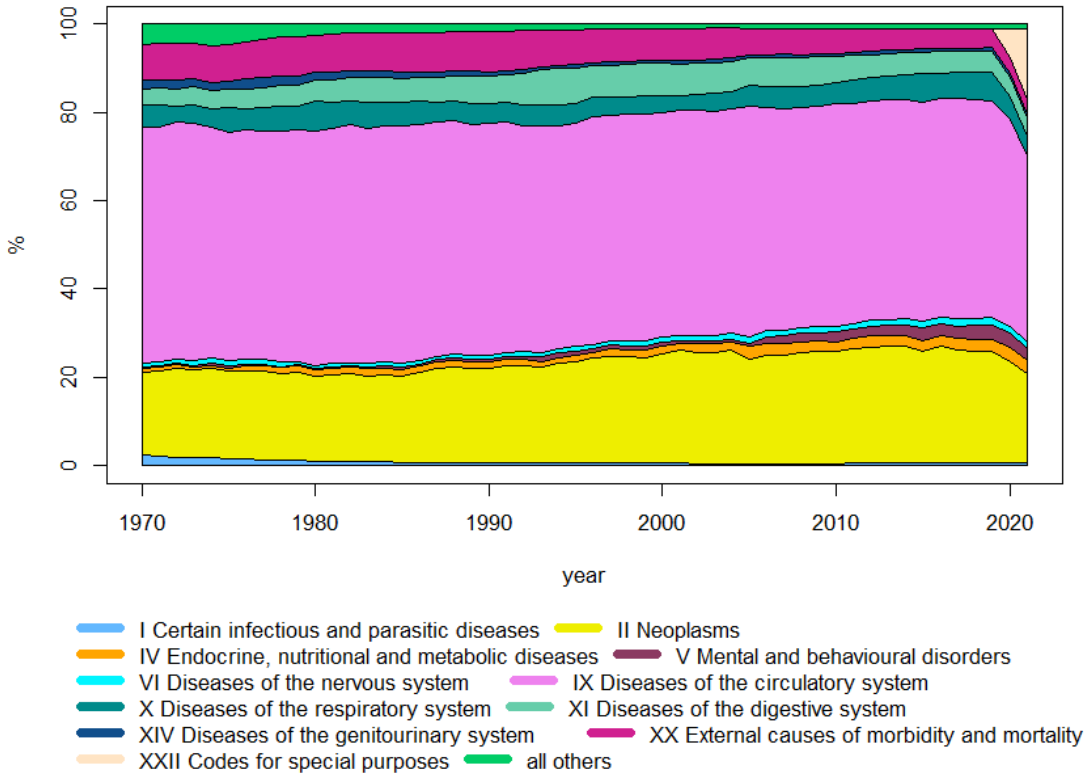
XX External causes of morbidity and mortality.[29]

In terms of the number of deaths, the group of mental and behavioural disorders could have been included in the analysis. However, as the fitted mortality models showed parameter uncertainty, we excluded this group of causes of death. The nine selected main causes accounted for approximately 95%, 89% and 80% of total deaths in 2019, 2020 and 2021, respectively. The decline in proportions was a consequence of the COVID-19 pandemic, during which time the XXII Codes for special purposes became more significant (see Figure 16).

Examining the distribution of all deaths by cause over time reveals that circulatory diseases and neoplasms have been the leading causes for decades. Since the early 1970s, a decline has been observed in the significance of certain infectious and parasitic diseases, diseases of the genitourinary system, and external causes within total mortality. On the other hand, the following causes of death have become more significant in recent decades: endocrine, nutritional and metabolic diseases, mental and behavioural disorders, and diseases of the nervous system. In 2020 and 2021, the proportions of the causes of death were significantly impacted by the COVID-19 epidemic. In Hungary, Tóth (2021a, 2022a, 2022b) estimated the impact of the COVID-19 pandemic on deaths and excess mortality.

[29] The group of mental and behavioural disorders was excluded from the analysis because subsequent model fits showed that the mortality trend analysis was uncertain.

Figure 16. Distribution of total deaths by primary cause between 1970 and 2021



Source: own calculation

The fitted parameters of the single-population models

Seven single-population models were selected for the mortality analysis by cause of death. The fitted parameter values are discussed in this section. Two of the seven mortality models examined – the Binomial Lee–Carter model and the Binomial Lee–Carter model with two factors – encountered robustness problems because they produced different results for certain causes of death compared to the Poisson versions of the same models. Thus, neither the Binomial Lee–Carter model nor the Binomial Lee–Carter model with two factors proved suitable in all cases. The following examples can be highlighted:

- the parameters of the Poisson and Binomial Lee–Carter models for diseases of the circulatory system for both sexes,
- the parameters of the Poisson and Binomial Lee–Carter models with two factors for the diseases of the nervous system, diseases of the circulatory system, diseases of the genitourinary system, and external causes of morbidity and mortality for males, and

- the parameters of the Poisson and Binomial Lee–Carter models with two factors for neoplasms and diseases of the circulatory system for females (see the online Appendix).

The figures for the fitted parameters of all the single-population models can be accessed in the online Appendix.

The parameters α_{xj} and β_{xj} represent the age-dependent effects by cause of death. The parameter κ_{tj} indicates the trend in cause-specific mortality. According to α_{xj}, the baseline shape of the log mortality curve, there are significant differences in mortality by cause of death. However, not all mortality models include this parameter. Based on the α_{xj} parameter, it is also possible to determine which of the nine main causes of death is more significant, and to analyze differences in mortality between sexes: the higher the values, the higher the mortality rates. Similar conclusions can be drawn regarding the ranking of causes of death for both men and women, as shown in Figure 5 based on the ASMR values.

Both positive and negative values can be observed for β_{xj} and κ_{tj}. These two parameters show whether mortality increased or decreased for certain causes of death at specific ages in the year under review. For most models and causes of death, the nonparametric β_{xj} takes positive values. According to the shapes of nonparametric β_{xj}, there are also differences according to the causes of death. The values are particularly high for the oldest women in the case of neoplasms, which corroborates the notion that this disease is less aggressive at an older age. Based on various models, the lowest β_{xj} values for diseases of the circulatory system are observed in the oldest age groups, suggesting that mortality improvement is faster among younger people. According to models by cause of death, similar conclusions can be drawn regarding which age groups show the greatest and which show the least improvement. The CBD and Plat models have parametric age effects, with values calculated for the 60–90 age range.

Mortality trends according to ASMR values (see Chapter III) correspond to κ_{tj} and, in particular, to the $\kappa_{tj}^{(1)}$ parameters. The parameter $\kappa_{tj}^{(1)}$, which describes the main mortality trend, remains consistent across different models that may incorporate multiple factors and/or parametric or nonparametric coefficients. The fitted parameters of the Poisson and Binomial Lee–Carter models, and the Poisson and Binomial Lee–Carter models with two factors (including the parameters κ_{tj}), demonstrate a high degree of similarity. Essentially, the two types of distribution only cause a slight difference in the

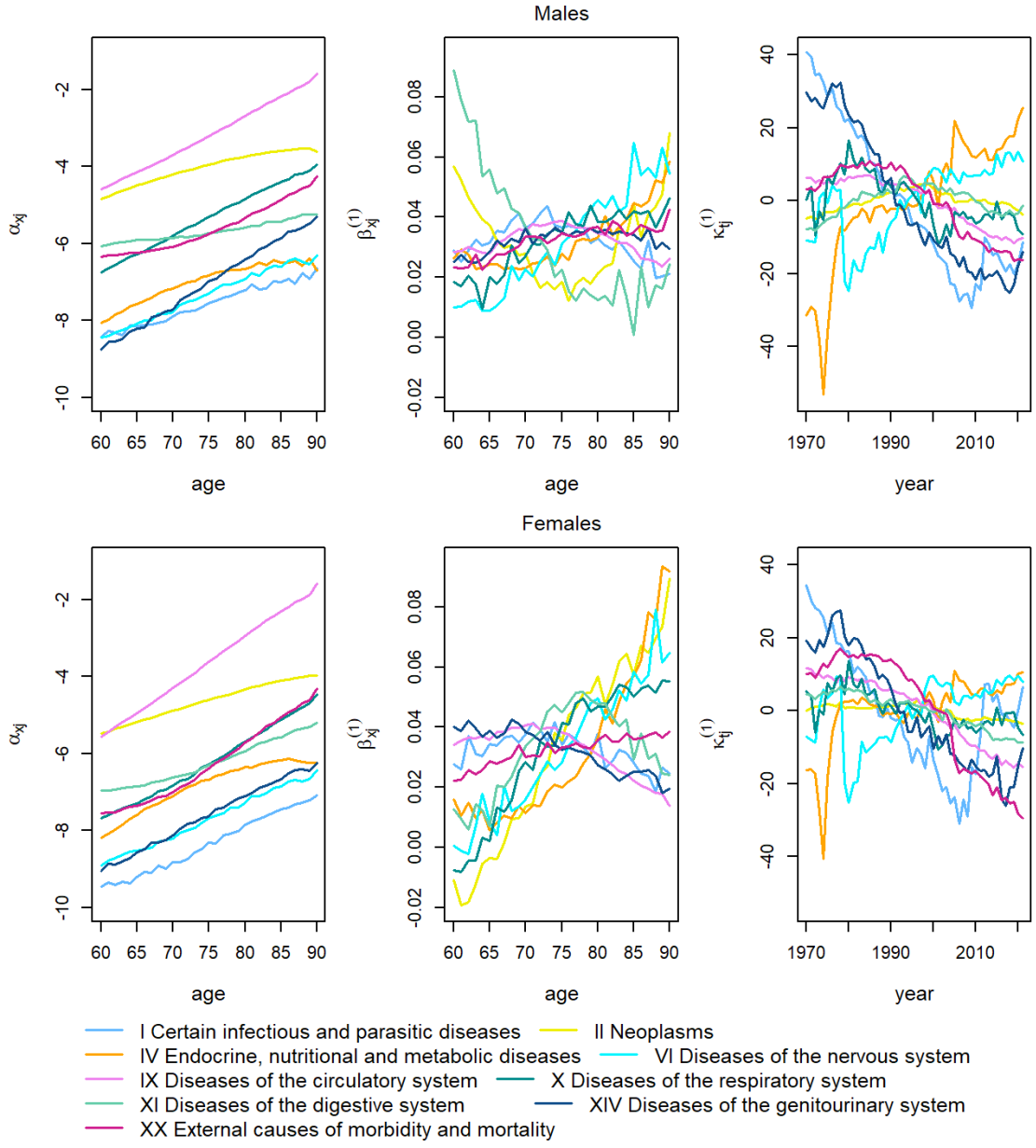
parameters. However, as previously mentioned, there were robustness problems relating to certain causes of death.

A decrease in $\kappa_{tj}^{(1)}$ values over time indicates an improvement in mortality. However, this is not the case for every cause of death. Mortality rates for certain infectious and parasitic diseases decreased over several decades, except in the 2010s. Neoplasms show a downward trend form the mid-1990s. Over the period, the mortality increased for endocrine, nutritional and metabolic diseases, as well as for diseases of the nervous system. Forecasting is more difficult for models with multiple factors because the fitted, time-dependent parameter values are more volatile. This is especially the case for the Poisson/Binomial Lee–Carter model with two factors and the reduced Plat model with three factors.[30] If there is no trend in the second or third indices, or if they are linear, there is no need for an extended model that includes multiple mortality indices. For this reason, none of the models were excluded from the analysis.

Our goal was to cluster the mortality indices of single-population models. Therefore, it is useful to compare the parameter trajectories fitted by cause of death. For example, the $\kappa_{tj}^{(1)}$ values for certain infectious and parasitic diseases and diseases of the genitourinary system are similar across the models, especially for men. This suggests that mortality trends for these two causes of death have been similar in recent decades. Therefore, it can be assumed that various factors, such as lifestyle changes and effective medicine, have played a similar role over time. Figures 17–19 show the fitted parameters of the Poisson Lee–Carter model and the Poisson Lee–Carter model with two factors by cause of death and sex. We selected these two models for presentation because one has the fewest stochastic parameters, while the other has the most. See the online Appendix for other models.

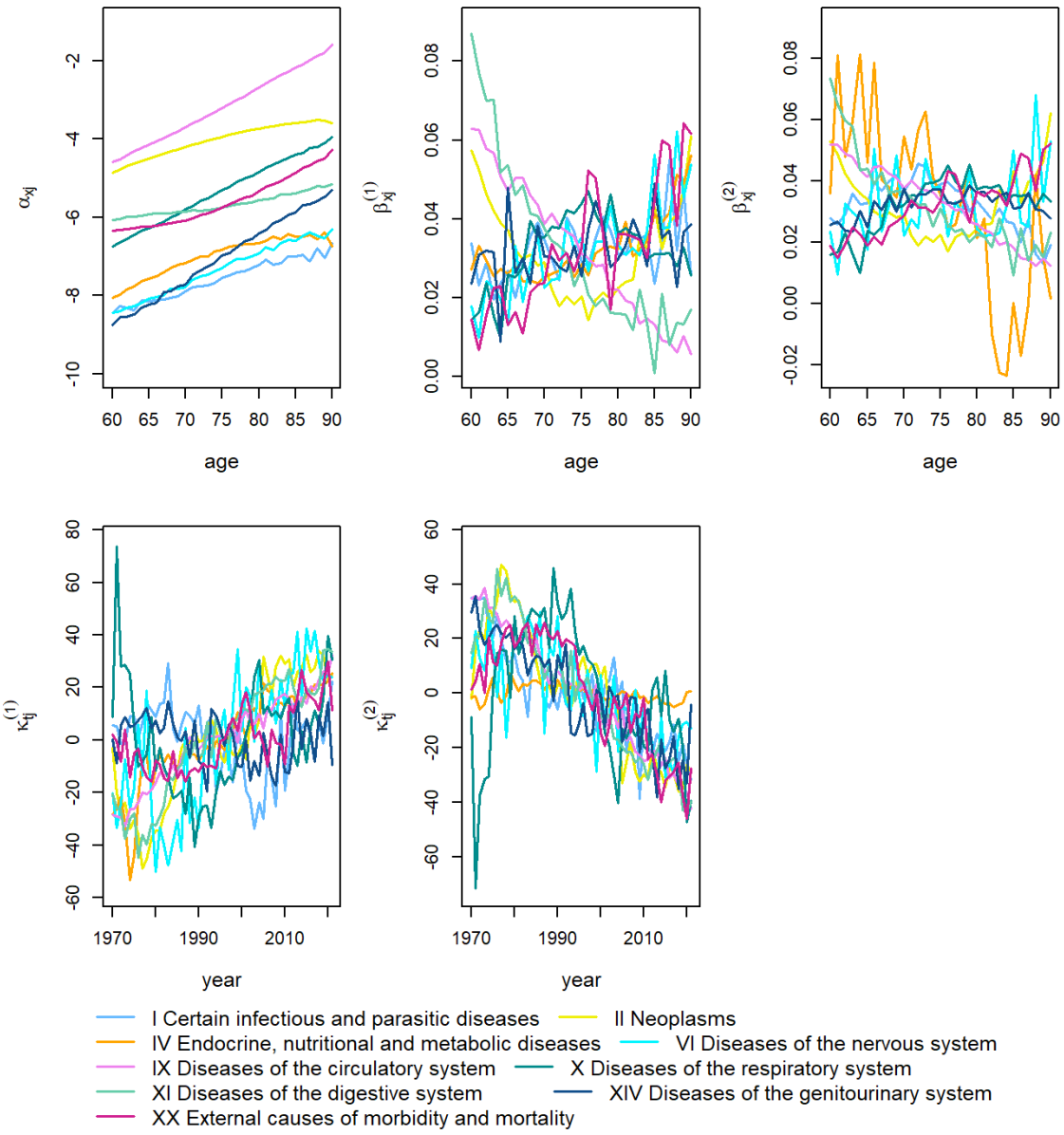
[30] The use of the latter model was not discarded. Only people aged 60 and over were studied.

Figure 17. The fitted parameters of the Poisson Lee–Carter model



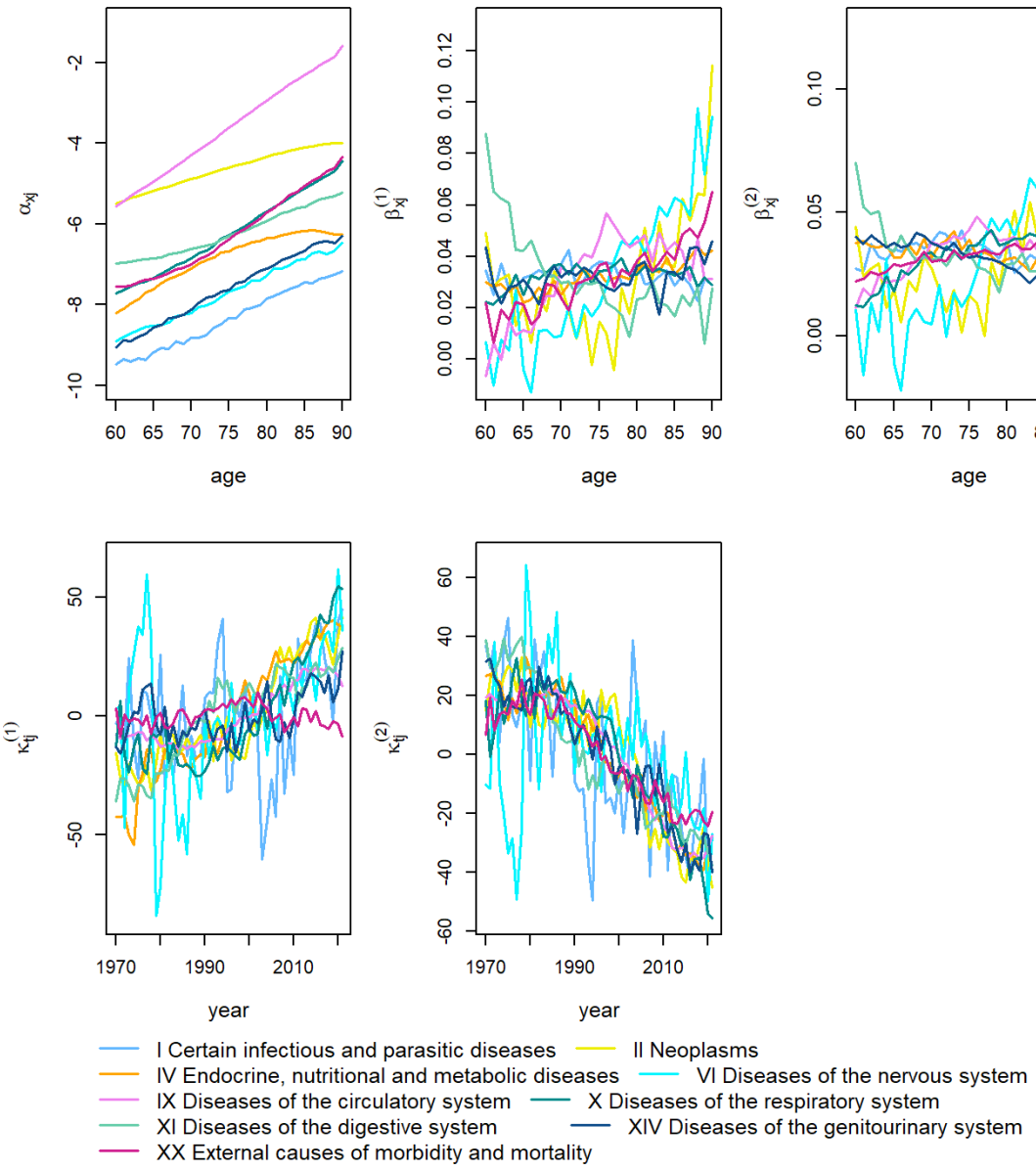
Source: own calculation

Figure 18. The fitted parameters of the Poisson Lee–Carter model with two factors for males



Source: own calculation

Figure 19. The fitted parameters of the Poisson Lee–Carter model with two factors for females



Source: own calculation

Creating multi-population versions of single-population models

The single-population models examined can easily be converted into multi-population models by setting one of the subpopulation-specific parameters to a value common to all subpopulations. At least one parameter in the initial equation of the multi-population model is shared by all subgroups. The seven selected mortality models are single-population models. For developing multi-population models, it is important that some form of age- or time-dependent similarity can be identified between the groups. A disadvantage of single-population models is that separate forecasts for different subpopulations can lead to divergence issues between groups with regard to mortality rates and, consequently, life expectancy. This can be avoided by using multi-population models. However, achieving coherence (see Li–Lee 2005) was not the primary objective of this research.[31] We constructed multi-population models based on the clustering of mortality indices and fitted a well-known model of this type: the Carter–Lee model. The other multi-population models can be considered innovations in this study.

Multi-population models enable us to examine characteristics that are specific to subgroups, as well as those that are common to the whole population. As part of our research, we created clusters of mortality indices according to the cause of death in each single-population model. This allowed us to examine the time-dependent relationship between different causes of death. Thus, the nonindependent relationship between causes of death can be captured. Using multi-population models reduces the number of parameters that need to be predicted, which is also considered an advantage. The crude (initial or central) mortality rates used in this research correspond to the observed probability of death. According to Zittersteyn–Alonso-García (2021), crude cause-of-death mortality is underestimated because it does not consider the possibility that an individual could have died from other causes. When other relevant causes of death are also taken into account, it is called net mortality. In our analysis, we focused on the primary cause of death, i.e., the cause deemed most responsible for death.

A variety of methods can be used to examine the nonindependent relationship between causes of death. One such method is copulas (see, e.g., Carriere 1994, Dimitrova et al. 2013, Kaishev et al. 2007, Li–Lu 2018, Zittersteyn–Alonso-García 2021). Cointegration analysis allows for the consideration of dependence between nonstationary

[31] Coherence can be measured: if the ratio of the death rates of different subpopulations converges to a constant over the long term, the model is coherent (see Li–Lee 2005).

time series, such as standardized age-specific mortality rates. To this end, Arnold–Sherris (2013, 2015) and Arnold–Glushko (2021) applied the vector error correction model (VECM). The rationale behind examining cointegration is that, while the significance of one cause of death increases, the significance of the other decreases. In addition to the use of copulas and VECM, multi-population models, which are the subject of the present study, may also be considered. These models may be appropriate for examining cause-specific crude death rates by jointly modeling causes of death that are similar in terms of mortality trends over time.

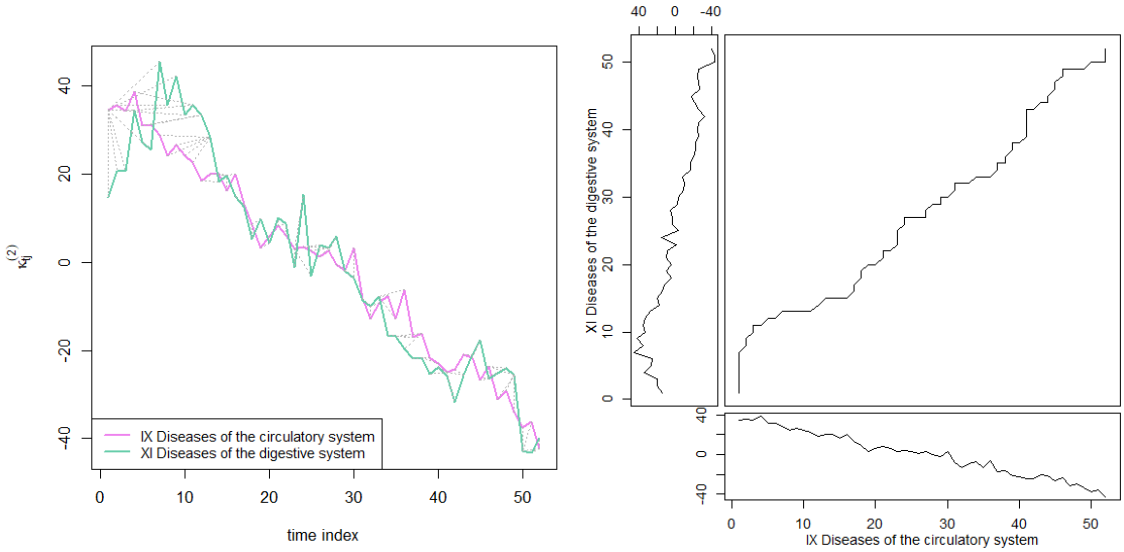
Examples of the use of clustering in relation to cause-specific mortality can be found in the literature. Li et al. (2019) employed cluster analysis to categorize various causes of death, thereby enhancing prediction accuracy. Piveteau (2020, 2021) presented the K-Lee–Carter model, which is inspired by the k-centroids clustering method. This model creates K different clusters from mortality time series, categorized by cause of death. In this study, we used two clustering methods, hierarchical clustering and k-medoids clustering, to categorize causes of death. Clustering the cause-specific mortality indices of the single-population models formed groups in which the indices were similar, but the groups themselves were clearly distinguishable from each other. We utilized the DTW algorithm to examine the similarity between the indices, as this is suitable for measuring the distance between time series.

The clusters of causes of death

Figures 20–21 illustrate the distances and optimal alignments between the chosen causes of death. These were selected as they demonstrate the greatest similarity. Additional figures on pairs of causes of death is available in the online Appendix, organized by mortality model. According to the Poisson Lee–Carter model with two factors, the smallest distance for males is between the second mortality indices for circulatory and digestive diseases (see Figure 20). The following two causes of death can be highlighted for females: endocrine, nutritional and metabolic diseases, and diseases of the circulatory system (see Figure 21). The centre boxes on the right-hand side of Figures 20–21 show the LCMs for the selected mortality indices, and demonstrate the optimal alignment.

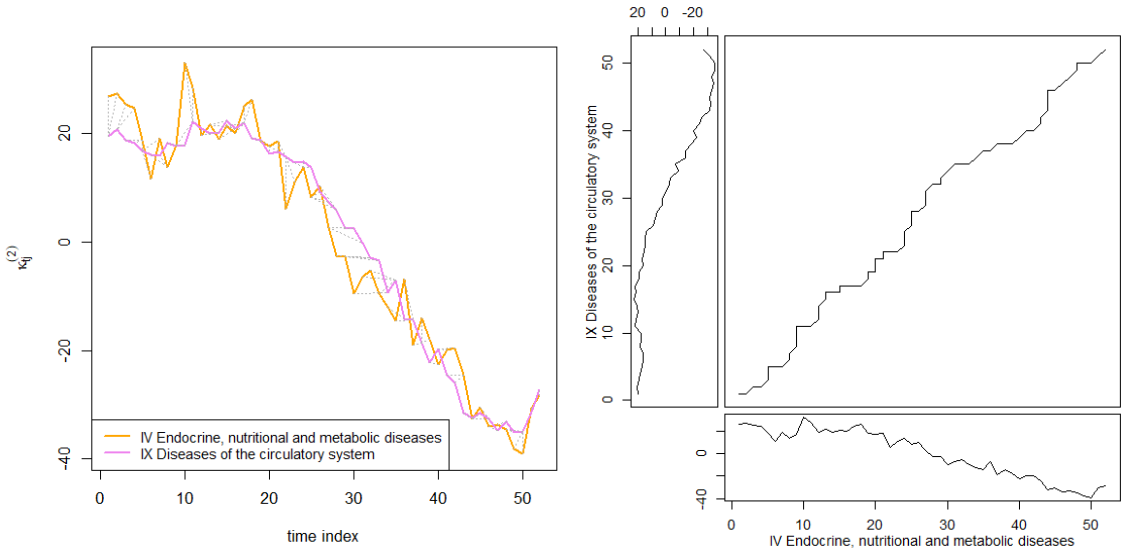
Tables in the online Appendix show distance values measured using the DTW algorithm for men and women separately.[32]

Figure 20. The optimal alignment between two mortality indices for males



Note: The series $\kappa_{tj}^{(2)}$ originate from the Poisson Lee–Carter model with two factors.
Source: own calculation

Figure 21. The optimal alignment between two mortality indices for females



Note: The series $\kappa_{tj}^{(2)}$ originate from the Poisson Lee–Carter model with two factors.
Source: own calculation

[32] The online Appendix only shows the distances between the time series that were later clustered.

Before clustering, the time series under investigation have to be standardized. The clustering was performed using two techniques: the hierarchical method and the k-medoids method (see Hastie et al. 2009). The latter method was implemented using partition around medoids (PAM) algorithm. The medoid is the prototype of a grouped time series: it is a member of a cluster that can be characterized by its main features. The ideal number of clusters was determined using dendrograms during hierarchical clustering (see figures in the online Appendix). Cluster validity indices such as the Silhouette, Dunn, COP, Davies–Bouldin, modified Davies–Bouldin, Calinski–Harabasz, and the score function were also useful for evaluating clustering (see Sardá-Espinosa 2017, 2019 and online Appendix). In addition, we repeated the clustering process using three different starting points to ensure the most appropriate outcome. It was not possible to cluster the mortality indices in the same way for all models. Clustering of the parameter $\kappa_{tj}^{(1)}$ proved to be appropriate for most models. The indices $\kappa_{tj}^{(2)}$ were categorized according to the Poisson/Binomial Lee–Carter model with two factors (see Tables 9–10). Figures 22–25 show the standardized values of the mortality indices by ideal number of clusters for the CBD model and the Binomial Lee–Carter model with two factors. The standardized mortality indices of other models can be found in the online Appendix.

Clustering was performed on a statistical basis according to the aforementioned criteria. However, we did not examine whether causes of death that grouped together due to similarities in mortality indices were attributable to the same factors (e.g., lifestyle or other external causes). We selected the most important causes of death for our research, but similar risk factors are not necessarily responsible for their development. We assumed that if similar mortality trends had been observed for certain causes of death in the past, these trends might also be observed in future. This common trend can be identified by fitting multi-population models.

Table 9. The ideal number of clusters and their members by model for males

Mortality indices of the single-population mortality models	Clusters					
	1	2	3	4	5	6
$\kappa_{tj}^{(1)}$ of Poisson Lee–Carter	I, XIV	II, XI	IX, XX	IV	VI	X
$\kappa_{tj}^{(1)}$ of Binomial Lee–Carter	I, XIV	II, XI	IX, XX	IV	VI	X
$\kappa_{tj}^{(2)}$ of Poisson Lee–Carter with two factors	I, XIV	II, IX, XI, XX	IV	VI	X	–
$\kappa_{tj}^{(2)}$ of Binomial Lee–Carter with two factors	I, XIV	II, IX, XI, XX	IV	VI	X	–
$\kappa_{tj}^{(1)}$ of CBD	I, XIV	II, XI	IX, XX	IV	VI	X
$\kappa_{tj}^{(1)}$ of reduced Plat with three factors	I, XIV	II, XI	IX, XX	IV	VI	X
$\kappa_{tj}^{(1)}$ of reduced Plat with two factors	I, XIV	II, XI	IX, XX	IV	VI	X

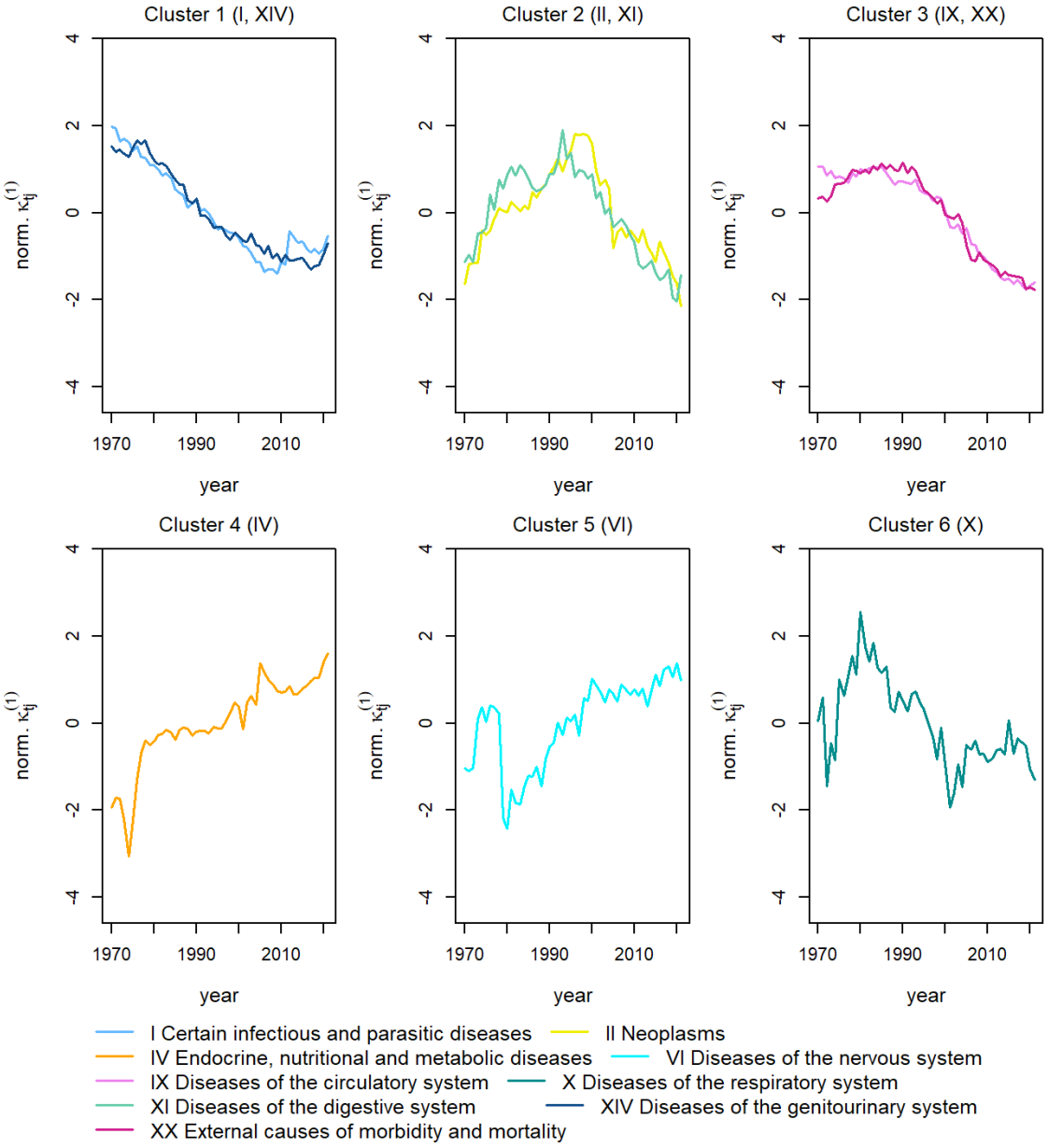
Note: The Roman numerals are used to sign the main groups of causes of death.
Source: own editing

Table 10. The ideal number of clusters and their members by model for females

Mortality indices of the single-population mortality models	Clusters					
	1	2	3	4	5	6
$\kappa_{tj}^{(1)}$ of Poisson Lee–Carter	IX, XI, XIV, XX	I	II	IV	VI	X
$\kappa_{tj}^{(1)}$ of Binomial Lee–Carter	IX, XI, XIV, XX	I	II	IV	VI	X
$\kappa_{tj}^{(2)}$ of Poisson Lee–Carter with two factors	IV, IX, XI, XIV, XX	I	II	VI	X	–
$\kappa_{tj}^{(2)}$ of Binomial Lee–Carter with two factors	IV, IX, XI, XIV, XX	I	II	VI	X	–
$\kappa_{tj}^{(1)}$ of CBD	IX, XI, XIV, XX	I	II	IV	VI	X
$\kappa_{tj}^{(1)}$ of reduced Plat with three factors	IX, XI, XIV, XX	I	II	IV	VI	X
$\kappa_{tj}^{(1)}$ of reduced Plat with two factors	IX, XI, XIV, XX	I	II	IV	VI	X

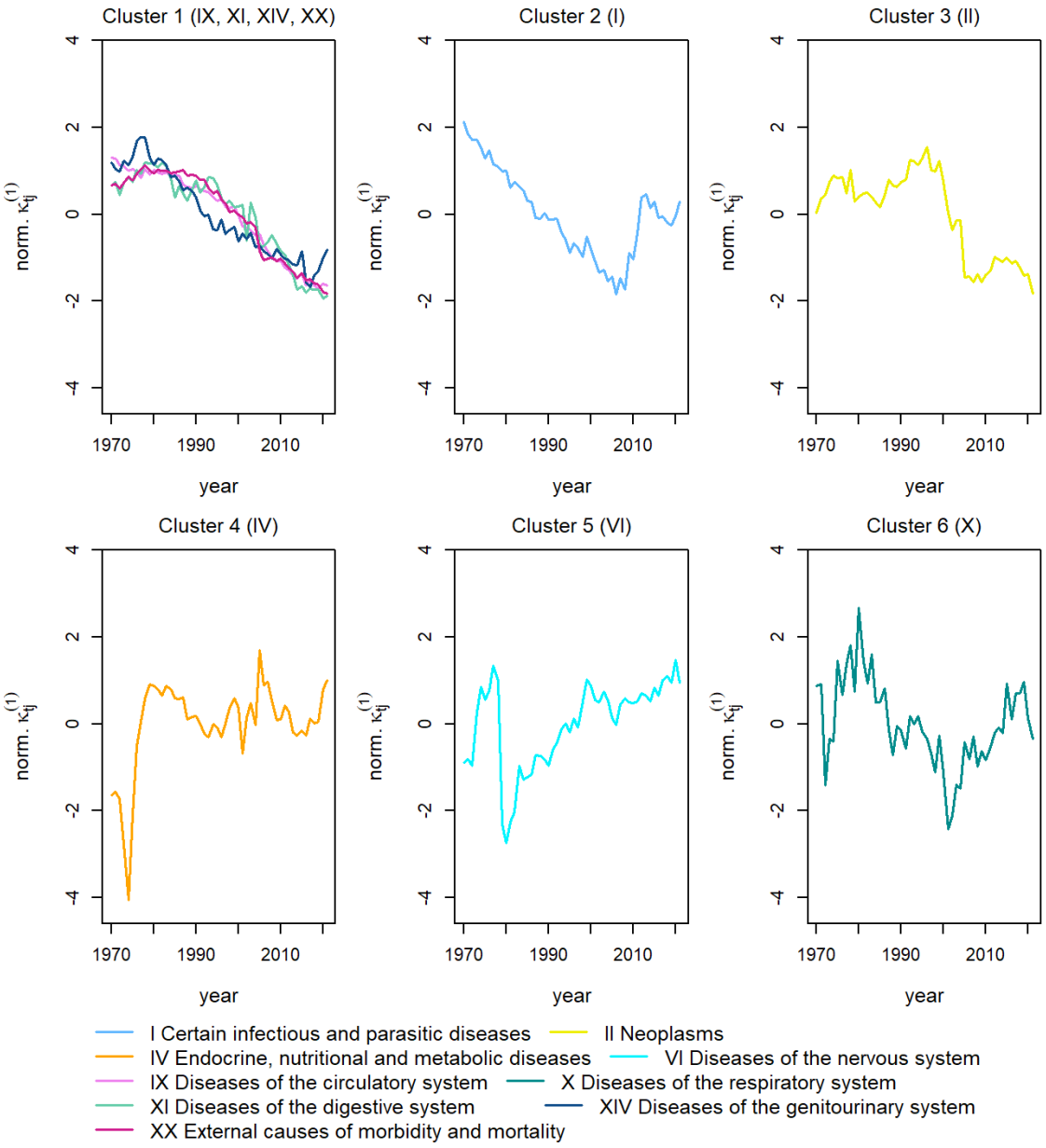
Note: The Roman numerals are used to sign the main groups of causes of death.
Source: own editing

Figure 22. The standardized mortality indices $\kappa_{tj}^{(1)}$ of the CBD model for males



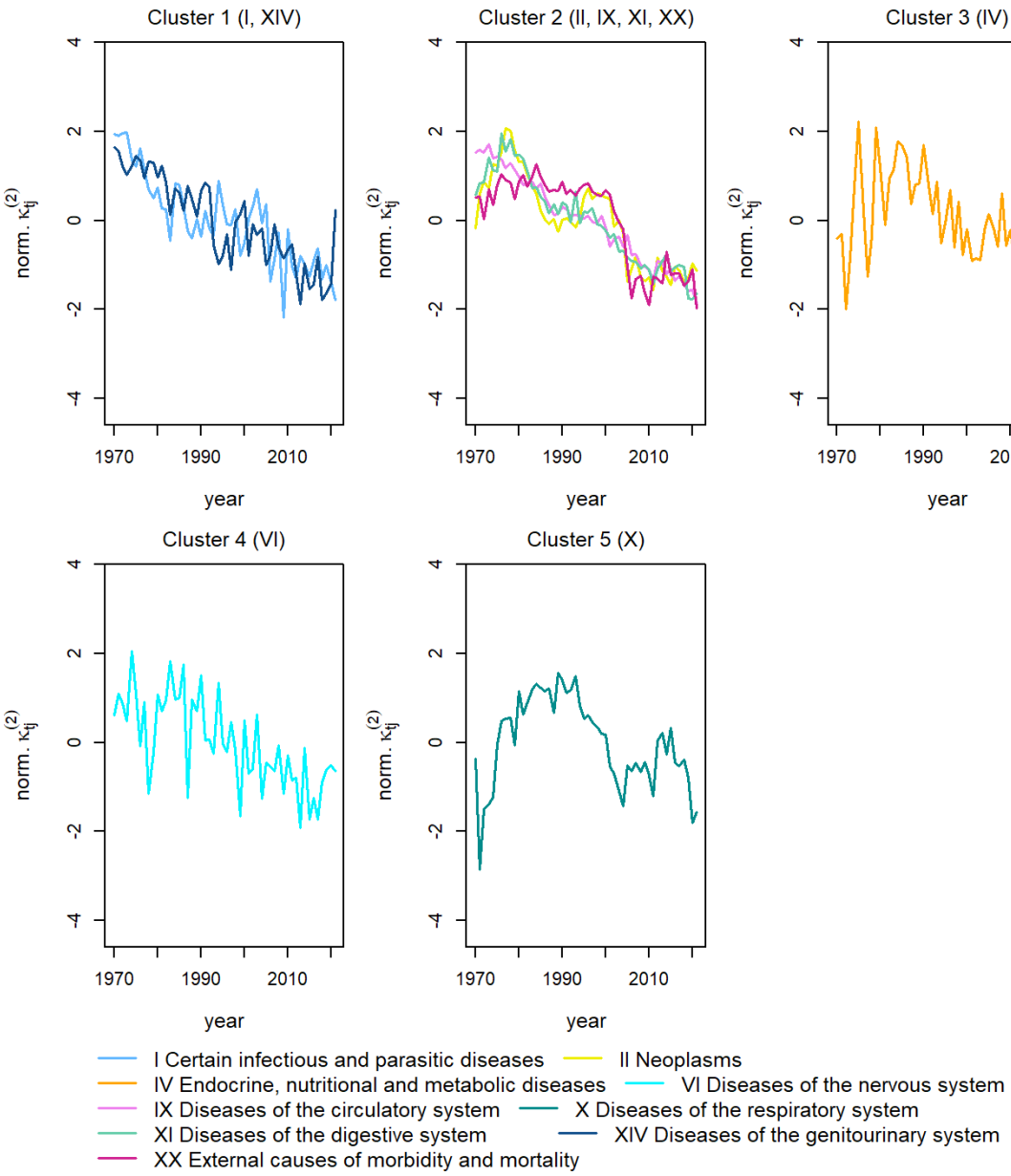
Source: own calculation

Figure 23. The standardized mortality indices $\kappa_{tj}^{(1)}$ of the CBD model for females



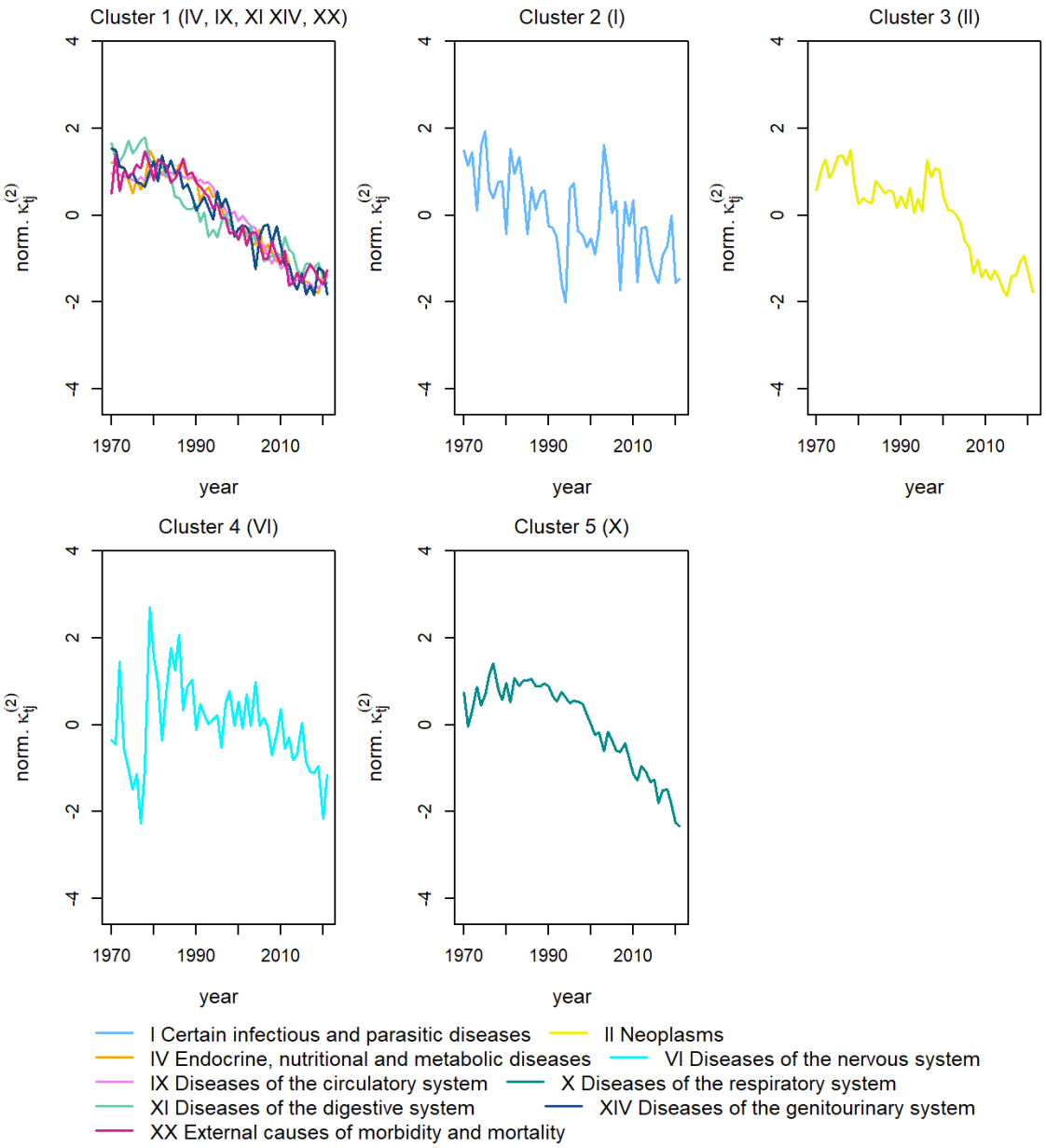
Source: own calculation

Figure 24. The standardized mortality indices $\kappa_{tj}^{(2)}$ of the Binomial Lee–Carter model with two factors for males



Source: own calculation

Figure 25. The standardized mortality indices $\kappa_{tj}^{(2)}$ of the Binomial Lee–Carter model with two factors for females



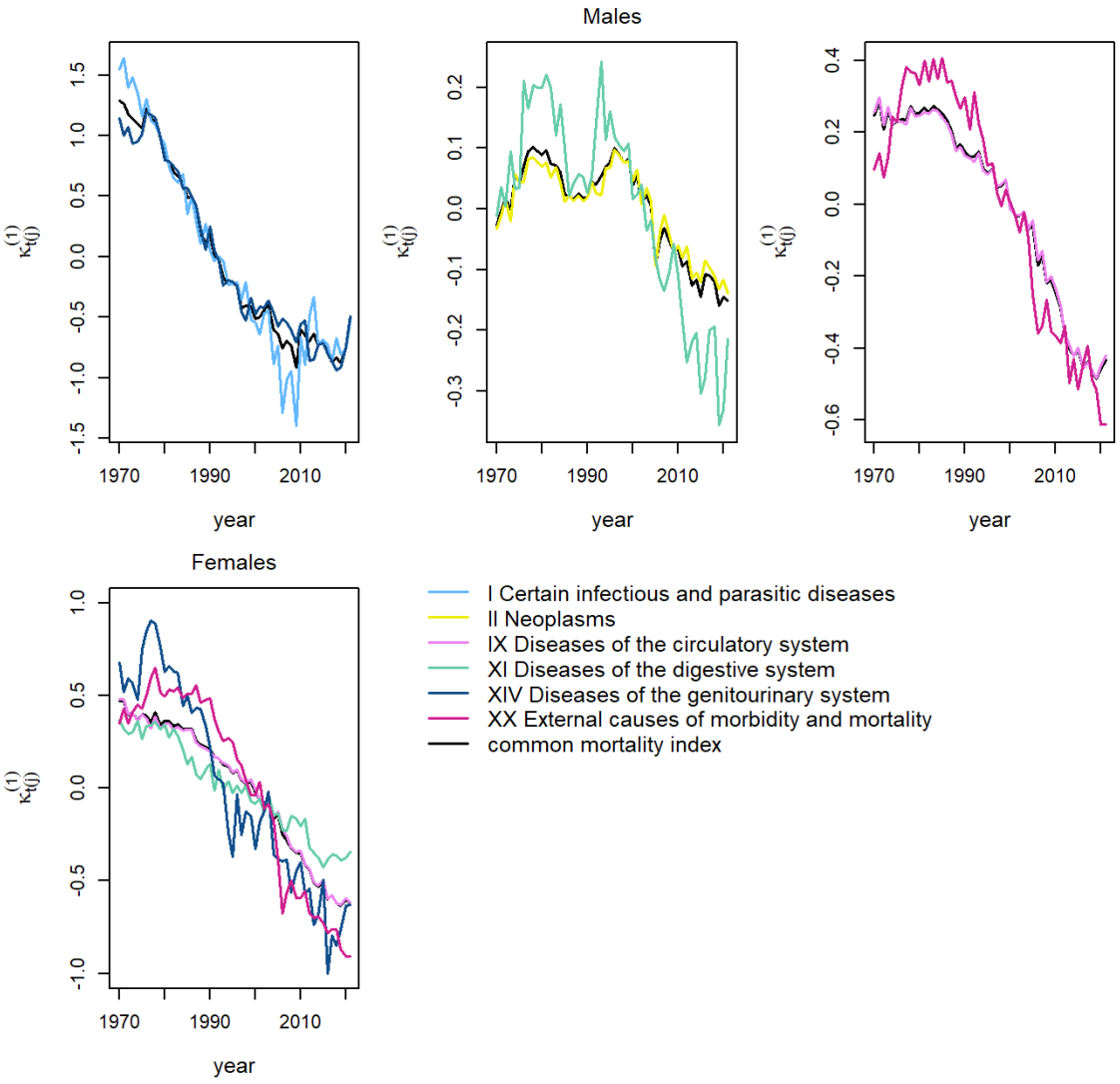
Source: own calculation

The fitted parameters of the multi-population models

In the case of causes of death grouped together, the common trend is represented by the common mortality index of multi-population models. However, this common trend is influenced to a greater extent by the cause of death that is responsible for a larger number of deaths and therefore carries greater weight. For example, the common mortality index of the reduced Plat model with three factors is more similar to the cause-specific index of circulatory diseases than to the trend in external causes of morbidity and mortality (see Figure 26). This could be a disadvantage of the multi-population models estimated using the maximum likelihood method. If the size differences between the main groups of causes of death are too significant, single-population models may be more appropriate for forecasting than multi-population ones. Alternatively, rather than using maximum likelihood, we can choose a method of estimation that analyzes mortality rates rather than the observed number of deaths and exposures directly. If the difference between the common mortality index of the multi-population model and the cause-specific mortality trend of the single-population models is small, it suggests that the global trend has been accurately captured for the clustered causes of death.

The multi-population variant of the Lee–Carter model with two factors has the greatest number of cause-specific parameters. This means that specific characteristics play the most important role in this multi-population mortality model for different causes of death. The reduced Plat model with three factors can be characterized by the most common parameters. When we compared the common mortality indices with the cause-specific mortality indices derived from the corresponding single-population models, we concluded that the multi-population version of the CBD model was not the most appropriate solution. The fitted parameters of all multi-population models by cause of death and sex can be found in the online Appendix. The following section presents the goodness of fit of mortality models based on five measures.

Figure 26. The common and cause-specific mortality indices $\kappa_{t(j)}^{(1)}$ of the reduced Plat model with three factors



Source: own calculation

The goodness of fit

The present study compared 14 single- and multi-population mortality models using both in-sample and out-of-sample analyses. For the in-sample analysis, we calculated the AIC and BIC values for the fitted models. The fitted mortality models include parsimonious models, as well as models with many parameters. The information criteria, AIC and BIC, represent a trade-off between goodness of fit and parsimony of the models. For example, the Poisson/Binomial Lee–Carter model with two factors has more parameters than the Poisson/Binomial Lee–Carter model. However, as shown in Tables 11–12, the former still has lower AIC and BIC values due to their favourable log-likelihood values. The online

Appendix presents the log-likelihood values of various mortality models by cause of death.

Tables 11–14 show the MAPE values resulting from a comparison of the fitted and observed crude mortality rates, categorized by cause of death and model. The reduced Plat model with three factors and the Poisson Lee–Carter model with two factors have the lowest MAPE values and are therefore the most favourable for both sexes. The multi-population versions of these models also performed well. The models mentioned earlier can also be highlighted based on the explanation ratio. According to LR statistics, the Binomial Lee–Carter model with two factors has the best performance. The multi-population version of this model also demonstrates better goodness of fit for a number of causes of death for both sexes in comparison to other models (see Tables 13–14).

The standardized residuals of the mortality models are plotted against age, calendar year, and year of birth on the heatmaps and scatter plots (see the online Appendix).[33] A random arrangement of the residuals is preferred. However, clustered patterns appear in the residuals for some cohort years in the Poisson/Binomial Lee–Carter model for six causes of death, which are denoted by Roman numerals as follows: II, IV, IX, X, XI, and XX. In most cases, the second or third additional index was able to improve the fit and compensate for the lack of age–period effects in models with multiple parameters compared to the simpler Lee–Carter model. However, the CBD model is an exception to this pattern.

Based on standardized residuals, the Poisson/Binomial Lee–Carter model with two factors and the reduced Plat models appear to capture the effects of age, year and cohort well, despite their initial equations not containing any additional cohort-dependent parameters. However, in two cases of death (II and IX), these models appear to be a poor fit because the residuals are not independent of age, calendar year and year of birth. For most single-population models, the values of standardized residuals are mostly between -2 and 2. Among the multi-population models, the extreme residuals (outside the ±2 range) suggest that the CBD model does not fit well for either men or women. The multi-population version of the Poisson Lee–Carter model with two factors appears to be an appropriate model due to the random arrangement of residuals.

The choice of nested models and the decision on whether to include additional factors are supported by the in-sample analysis measures (information criteria, MAPE

[33] We used the *StMoMo* package of Villegas et al. (2018) to prepare heatmaps and scatter plots.

values, explanation ratios, LR statistics, and standardized residuals), as well as by the shape of the mortality indices. It is important to avoid over-parameterization and strive for parsimonious model building. There is no need to expand the model if the values of the various statistics do not improve significantly as a result of the added factors. Residual plots confirmed that adding a second factor improves the goodness of fit in the case of the Lee–Carter model (both single- and multi-population versions).

Overall, single-population models produced a better fit than the multi-population versions. However, for most causes of death (e.g., groups I, II, XI, XIV and XX for males and groups IV, XIV and XX for females), there was no significant difference between the single-population and multi-population versions of the Binomial Lee–Carter model with two factors. The multi-population extension of the CBD model performed the worst. This is mainly supported by the MAPE and the explanatory ratio. The Binomial Lee–Carter models with one or two factors and their multi-population extensions produced the most favourable results in the in-sample analysis (see Tables 11–14).

Table 11. The single-population mortality models' goodness of fit for males, 1970–2021

Causes of death	Measures	Poisson Lee–Carter	Binomial Lee–Carter	Poisson Lee–Carter with two factors	Binomial Lee–Carter with two factors	CBD	reduced Plat with three factors	reduced Plat with two factors
I	AIC	361,644	*361,440*	361,633	**361,420**	361,679	361,664	361,839
	BIC	362,248	**362,043**	362,673	362,460	*362,239*	362,655	362,555
	MAPE	0.322	0.325	**0.302**	*0.305*	0.354	0.306	0.336
	R	0.796	0.793	**0.814**	*0.811*	0.768	*0.811*	0.778
	LR	892,559	*892,156*	892,386	**891,975**	892,411	892,435	892,712
II	AIC	6,656,578	*6,640,008*	6,655,230	**6,638,481**	6,641,909	6,655,221	6,655,886
	BIC	6,657,181	*6,640,611*	6,656,270	**6,639,521**	6,642,469	6,656,212	6,656,602
	MAPE	0.065	0.073	**0.050**	*0.058*	0.097	**0.050**	0.062
	R	0.510	0.445	*0.683*	0.612	n. a.	**0.685**	0.500
	LR	26,543,064	*26,531,324*	26,541,555	**26,529,667**	26,533,189	26,541,564	26,542,331
IV	AIC	600,809	*600,397*	600,764	**600,359**	600,802	600,758	600,812
	BIC	601,412	**601,000**	601,804	601,399	*601,362*	601,748	601,528
	MAPE	0.248	0.250	*0.231*	0.234	0.304	**0.230**	0.251
	R	0.720	0.719	*0.758*	0.753	0.668	**0.768**	0.746
	LR	1,568,020	*1,567,239*	1,567,813	**1,567,040**	1,567,659	1,567,825	1,567,981
VI	AIC	402,895	*402,602*	402,887	**402,597**	402,612	402,888	402,855
	BIC	403,498	*403,206*	403,926	403,637	**403,172**	403,879	403,571
	MAPE	0.269	0.270	**0.254**	0.256	0.277	*0.255*	0.261
	R	0.569	0.568	**0.607**	*0.603*	0.562	*0.603*	0.590
	LR	989,352	*988,780*	989,182	**988,613**	988,805	989,201	989,270
IX	AIC	11,528,228	*11,420,878*	11,525,821	**11,418,113**	11,423,418	11,525,366	11,526,737
	BIC	11,528,831	*11,421,481*	11,526,860	**11,419,152**	11,423,978	11,526,357	11,527,453
	MAPE	0.052	0.062	*0.038*	0.048	0.066	**0.036**	0.045
	R	0.906	0.864	*0.951*	0.910	0.838	**0.956**	0.931
	LR	55,831,294	*55,799,623*	55,828,726	**55,796,851**	55,801,831	55,828,289	55,829,762
X	AIC	2,068,078	2,065,313	2,067,844	**2,065,065**	*2,065,179*	2,067,662	2,067,920
	BIC	2,068,681	*2,065,917*	2,068,883	2,066,104	**2,065,739**	2,068,652	2,068,636
	MAPE	0.102	0.107	*0.092*	0.097	0.106	**0.088**	0.097
	R	0.753	0.731	*0.797*	0.773	0.742	**0.817**	0.778
	LR	6,577,797	6,574,354	6,577,401	**6,573,947**	*6,574,235*	6,577,237	6,577,597
XI	AIC	1,769,622	1,768,257	1,768,170	**1,766,785**	*1,767,304*	1,768,214	1,768,376
	BIC	1,770,225	1,768,860	1,769,210	**1,767,825**	*1,767,865*	1,769,204	1,769,093
	MAPE	0.151	0.158	**0.108**	0.115	0.133	*0.110*	0.121
	R	0.392	0.379	**0.683**	0.659	0.525	*0.674*	0.592
	LR	5,392,455	5,390,278	5,390,842	**5,388,651**	*5,389,345*	5,390,903	5,391,168
XIV	AIC	523,007	*522,522*	523,011	**522,517**	522,632	523,061	523,079
	BIC	523,610	**523,125**	524,050	523,556	*523,193*	524,052	523,795
	MAPE	0.239	0.243	**0.226**	*0.229*	0.249	*0.229*	0.237
	R	0.850	0.845	**0.863**	*0.859*	0.840	0.858	0.850
	LR	1,364,864	*1,364,003*	1,364,706	**1,363,837**	1,364,129	1,364,774	1,364,894
XX	AIC	1,683,617	*1,682,030*	1,683,585	**1,681,982**	1,684,459	1,683,574	1,683,583
	BIC	1,684,220	**1,682,633**	1,684,624	*1,683,022*	1,685,019	1,684,565	1,684,299
	MAPE	0.103	0.109	**0.096**	0.103	0.156	**0.096**	*0.100*
	R	0.848	0.831	**0.866**	0.847	0.662	**0.866**	*0.853*
	LR	5,101,376	*5,099,063*	5,101,182	**5,098,855**	5,101,495	5,101,189	5,101,300

Notes: The best values are highlighted in bold, and the second-best values are in italics. The abbreviation "n.a." refers to invaluable values, i.e., values that are not available, and indicates a poor fit.
Source: own calculation

Table 12. The single-population mortality models' goodness of fit for females, 1970–2021

Causes of death	Mea-sures	Poisson Lee–Carter	Binomial Lee–Carter	Poisson Lee–Carter with two factors	Binomial Lee–Carter with two factors	CBD	reduced Plat with three factors	reduced Plat with two factors
I	AIC	272,499	*272,273*	272,485	**272,264**	272,311	272,519	272,540
	BIC	273,102	*272,876*	273,525	273,304	**272,871**	273,510	273,257
	MAPE	0.385	0.384	**0.360**	*0.361*	0.385	0.365	0.380
	R	0.652	0.653	**0.680**	**0.680**	0.647	*0.676*	0.655
	LR	630,942	*630,484*	630,766	**630,313**	630,538	630,817	630,941
II	AIC	6,295,225	*6,283,114*	6,295,033	**6,282,863**	6,283,730	6,294,916	6,295,084
	BIC	6,295,828	**6,283,717**	6,296,072	*6,283,903*	6,284,290	6,295,907	6,295,800
	MAPE	0.055	0.065	*0.049*	0.060	0.071	**0.048**	0.052
	R	0.569	0.470	*0.643*	0.538	0.347	**0.659**	0.596
	LR	23,454,457	*23,440,524*	23,454,104	**23,440,115**	23,441,143	23,454,005	23,454,275
IV	AIC	1,048,594	1,047,648	1,047,034	**1,046,119**	1,047,721	*1,047,012*	1,047,208
	BIC	1,049,197	1,048,251	1,048,073	**1,047,158**	1,048,281	1,048,003	*1,047,924*
	MAPE	0.215	0.220	**0.151**	0.157	0.259	*0.152*	0.172
	R	0.543	0.537	**0.783**	0.769	0.482	*0.778*	0.721
	LR	2,900,318	*2,898,538*	2,898,596	**2,896,849**	2,898,625	2,898,592	2,898,890
VI	AIC	490,472	490,029	490,435	**489,991**	*490,011*	490,455	490,437
	BIC	491,075	*490,632*	491,475	491,030	**490,571**	491,445	491,154
	MAPE	0.243	0.247	**0.225**	*0.230*	0.248	0.232	0.239
	R	0.598	0.591	**0.639**	*0.629*	0.586	0.625	0.604
	LR	1,226,981	1,226,080	1,226,782	**1,225,881**	*1,226,079*	1,226,819	1,226,904
IX	AIC	14,513,216	*14,334,420*	14,512,488	**14,333,377**	14,339,684	14,512,248	14,514,249
	BIC	14,513,819	*14,335,023*	14,513,527	**14,334,417**	14,340,244	14,513,239	14,514,965
	MAPE	0.043	0.054	*0.036*	0.049	0.068	**0.035**	0.050
	R	0.966	0.937	*0.975*	0.947	0.893	**0.976**	0.944
	LR	74,494,977	*74,427,177*	74,494,087	**74,426,054**	74,431,800	74,493,865	74,495,968
X	AIC	1,679,297	1,676,811	1,677,928	**1,675,431**	*1,676,004*	1,677,758	1,677,988
	BIC	1,679,900	1,677,414	1,678,968	**1,676,471**	*1,676,564*	1,678,749	1,678,704
	MAPE	0.147	0.152	*0.105*	0.110	0.127	**0.099**	0.110
	R	0.528	0.515	*0.772*	0.753	0.662	**0.794**	0.744
	LR	5,066,020	5,061,926	5,064,489	**5,060,386**	*5,061,134*	5,064,337	5,064,669
XI	AIC	1,613,122	*1,611,458*	1,612,556	**1,610,867**	1,612,073	1,612,535	1,612,998
	BIC	1,613,725	*1,612,061*	1,613,596	**1,611,907**	1,612,633	1,613,526	1,613,715
	MAPE	0.120	0.126	*0.101*	0.107	0.142	**0.100**	0.120
	R	0.539	0.514	*0.682*	0.653	0.415	**0.687**	0.549
	LR	4,692,248	*4,689,175*	4,691,520	**4,688,424**	4,689,804	4,691,517	4,692,082
XIV	AIC	516,953	*516,486*	516,851	**516,383**	**516,383**	516,850	516,827
	BIC	517,556	*517,090*	517,890	517,422	**516,943**	517,841	517,544
	MAPE	0.255	0.256	*0.235*	0.237	0.252	**0.233**	0.240
	R	0.778	0.776	*0.802*	0.799	0.781	**0.805**	0.794
	LR	1,298,844	1,297,929	1,298,580	**1,297,664**	*1,297,842*	1,298,597	1,298,676

XX	AIC	1,669,607	*1,666,738*	1,669,513	**1,666,641**	1,668,549	1,669,521	1,669,507
	BIC	1,670,210	**1,667,341**	1,670,552	*1,667,680*	1,669,109	1,670,512	1,670,223
	MAPE	0.114	0.119	**0.108**	0.114	0.161	**0.108**	*0.111*
	R	0.927	0.920	**0.934**	0.926	0.835	*0.933*	0.930
	LR	5,104,730	*5,100,202*	5,104,474	**5,099,946**	5,102,024	5,104,501	5,104,589

Note: The best values are highlighted in bold, and the second-best values are in italics.
Source: own calculation

Table 13. The multi-population mortality models' goodness of fit for males, 1970–2021

Causes of death	Mea-sures	Poisson Carter–Lee	Binomial Carter–Lee	Poisson Lee–Carter with two factors (multi-population extension)	Binomial Lee–Carter with two factors (multi-population extension)	CBD (multi-population extension)	reduced Plat with three factors (multi-population extension)	reduced Plat with two factors (multi-population extension)
I	AIC	361,954	361,753	*361,716*	**361,506**	391,101	361,838	362,110
	BIC	362,557	**362,356**	362,755	*362,545*	391,661	362,829	362,826
	MAPE	0.349	0.351	**0.308**	*0.311*	n. a.	0.328	0.366
	R	0.774	0.772	**0.807**	*0.804*	n. a.	0.793	0.754
	LR	892,869	*892,469*	*892,469*	**892,102**	n. a.	892,609	892,982
II	AIC	6,656,708	*6,640,142*	6,655,448	**6,639,007**	6,660,621	6,655,258	6,656,022
	BIC	6,657,311	*6,640,745*	6,656,488	**6,640,047**	6,661,181	6,656,249	6,656,739
	MAPE	0.067	0.074	*0.053*	0.062	n. a.	**0.050**	0.064
	R	0.487	0.429	*0.654*	0.560	n. a.	**0.682**	0.462
	LR	26,543,195	**26,531,457**	26,541,773	26,552,196	n. a.	*26,541,600*	26,542,467
IX	AIC	11,528,331	**11,421,015**	11,525,976	*11,427,668*	11,434,935	11,525,395	11,526,824
	BIC	11,528,934	**11,421,619**	11,527,015	*11,428,707*	11,435,495	11,526,386	11,527,540
	MAPE	0.053	0.062	*0.039*	0.075	0.110	**0.036**	0.046
	R	0.905	0.862	*0.948*	0.787	0.705	**0.956**	0.929
	LR	55,831,398	*55,799,743*	55,828,881	55,967,127	**55,562,077**	55,828,318	55,829,849
XI	AIC	1,770,240	1,768,827	*1,768,284*	**1,766,903**	n. a.	1,768,468	1,769,061
	BIC	1,770,843	1,769,430	*1,769,324*	**1,767,943**	n. a.	1,769,459	1,769,777
	MAPE	0.163	0.168	**0.113**	0.119	n. a.	*0.117*	0.138
	R	0.334	0.327	**0.658**	*0.637*	n. a.	0.632	0.486
	LR	5,393,074	*5,390,847*	5,390,956	**5,389,614**	n. a.	5,391,158	5,391,853
XIV	AIC	523,233	*522,753*	523,080	**522,589**	532,854	523,159	523,290
	BIC	523,836	**523,356**	524,119	*523,628*	533,414	524,149	524,006
	MAPE	0.245	0.250	*0.233*	0.237	0.811	**0.232**	0.244
	R	0.841	0.835	**0.854**	*0.850*	0.299	**0.854**	0.840
	LR	1,365,090	1,364,233	1,364,775	*1,364,081*	**1,330,674**	1,364,872	1,365,105
XX	AIC	1,684,938	*1,683,131*	1,683,730	**1,682,135**	n. a.	1,683,944	1,684,836
	BIC	1,685,541	*1,683,734*	1,684,769	**1,683,174**	n. a.	1,684,935	1,685,552
	MAPE	0.139	0.139	**0.101**	*0.106*	n. a.	0.108	0.136
	R	0.748	0.749	**0.852**	0.836	n. a.	*0.837*	0.759
	LR	5,102,697	*5,100,160*	5,101,327	**5,100,121**	n. a.	5,101,559	5,102,553

Notes: The best values are highlighted in bold, and the second-best values are in italics. The abbreviation "n.a." refers to invaluable values, i.e., values that are not available, and indicates a poor fit.
Source: own calculation

Table 14. The multi-population mortality models' goodness of fit for females, 1970–2021

Causes of death	Mea-sures	Poisson Carter–Lee	Binomial Carter–Lee	Poisson Lee–Carter with two factors (multi-population extension)	Binomial Lee–Carter with two factors (multi-population extension)	CBD (multi-population extension)	reduced Plat with three factors (multi-population extension)	reduced Plat with two factors (multi-population extension)
IV	AIC	–	–	*1,047,224*	**1,046,270**	–	–	–
	BIC	–	–	*1,048,263*	**1,047,310**	–	–	–
	MAPE	–	–	**0.160**	*0.163*	–	–	–
	R	–	–	**0.762**	*0.752*	–	–	–
	LR	–	–	*2,898,786*	**2,897,230**	–	–	–
IX	AIC	14,513,327	**14,334,550**	14,512,562	*14,350,773*	14,380,007	14,512,315	14,514,362
	BIC	14,513,930	**14,335,153**	14,513,601	*14,351,812*	14,380,567	14,513,305	14,515,078
	MAPE	0.043	0.054	*0.036*	0.074	0.188	**0.035**	0.050
	R	0.966	0.937	*0.975*	0.895	0.632	**0.976**	0.944
	LR	74,495,088	*74,427,291*	74,494,161	74,690,656	**73,987,029**	74,493,932	74,496,081
XI	AIC	1,615,887	1,614,779	*1,612,704*	**1,611,016**	n. a.	1,613,040	1,615,499
	BIC	1,616,490	1,615,382	*1,613,744*	**1,612,055**	n. a.	1,614,031	1,616,215
	MAPE	n. a.	n. a.	**0.104**	*0.111*	n. a.	0.112	n. a.
	R	n. a.	n. a.	**0.654**	*0.630*	n. a.	0.598	n. a.
	LR	4,695,013	4,692,489	*4,691,668*	**4,689,095**	n. a.	4,692,022	4,694,582
XIV	AIC	519,158	518,527	*516,966*	**516,480**	n. a.	517,378	519,030
	BIC	519,761	519,130	*518,005*	**517,519**	n. a.	518,369	519,746
	MAPE	0.402	0.391	**0.245**	*0.246*	n. a.	0.267	0.391
	R	0.560	0.578	**0.788**	**0.788**	n. a.	*0.760*	0.577
	LR	1,301,049	1,299,967	*1,298,695*	**1,297,819**	n. a.	1,299,125	1,300,879
XX	AIC	1,675,790	1,672,390	*1,669,698*	**1,666,804**	n. a.	1,670,474	1,673,488
	BIC	1,676,393	1,672,994	*1,670,737*	**1,667,843**	n. a.	1,671,465	1,674,204
	MAPE	0.247	0.240	**0.112**	*0.117*	n. a.	0.137	0.234
	R	0.739	0.751	**0.928**	*0.922*	n. a.	0.898	0.743
	LR	5,110,913	5,105,826	*5,104,660*	**5,101,544**	n. a.	5,105,454	5,108,570

Notes: The best values are highlighted in bold, and the second-best values are in italics. The abbreviation "n.a." refers to invaluable values, i.e., values that are not available, and indicates a poor fit. The sign "–" denotes that there is no model for the cause of death that has been studied.
Source: own calculation

Forecasting the mortality indices

Following the in-sample analysis, cause-specific mortality was projected, for which the mortality indices had to be predicted. Box plots were used to identify any outliers in the time series. In terms of the forecast time horizon, we assumed that no unexpected events (e.g., another wave of the COVID-19 pandemic) would occur that would require the inclusion of a dummy variable in the time series model. Where necessary, the impact of the COVID-19 pandemic was filtered out in an additive way with regard to the critical

years of the base period. The online Appendix contains more detailed information on which years should be treated as outliers.

In models with multiple mortality indices, we assumed that the indices were independent of each other. In addition, we assumed that the cause-specific mortality indices were independent of each other. This means that single-population models predicted cause-specific mortality without considering trends in other causes of death. In multi-population models considering grouped causes of death, we took the possibility of correlation into account, given that the models contained a common mortality index due to similar cause-specific trends. We also used ARIMA models to forecast cause-specific and common mortality indices separately. Multivariate random walk with drift processes, vector autoregressive (VAR) models, and VECMs may be considered as alternatives to ARIMA models, as they can capture relationships between mortality indices (especially in the case of single-population models).[34] With regard to common mortality indices in multi-population models, we assumed that the current trend of clustered causes of death would continue in the future.

The selection of the appropriate ARIMA processes was based on the Box–Jenkins methodology (see, e.g., Enders 2014). The ADF and KPSS unit root tests were used to check stationarity of the mortality indices. For each cause-specific time series, a first differencing was necessary. Different ARIMA(p,d,q) models were fitted for each mortality index with the range p, d, q = 0, 1, 2. The normality and autocorrelation of the ARIMA model residuals were tested using various tests, including the Shapiro–Wilk test, the Ljung–Box Q-statistic and the Breusch–Godfrey test. These tests and the information criteria (AIC and BIC)[35] were used to select the most appropriate ARIMA models (see Table 15).

Furthermore, an out-of-sample analysis was conducted for the mortality indices of the single-population models. The base period was shortened to between 1970 and 2015, and the forecast was made for the period between 2016 and 2019.[36] The MAPE measure was calculated to compare the fitted and predicted values. This out-of-sample analysis confirmed the selection of the most appropriate ARIMA models. The results by

[34] Jarner and Jallbjørn (2020), for example, analyzed cointegration within the Lee–Carter framework. We also considered using the VAR model and the VECM for forecasting in our research, but discarded these models due to their inadequate results.

[35] The following formulas are used: $AIC = T\ ln(sum\ of\ squared\ residuals) + 2n$ and $BIC = T\ ln(sum\ of\ squared\ residuals) + n\ ln(T)$ where n is the number of estimated parameters (p + q + possible constant term), and T is the number of usable observations (see, e.g., Enders 2014).

[36] The years 2020 and 2021 have been omitted because of the COVID-19 pandemic.

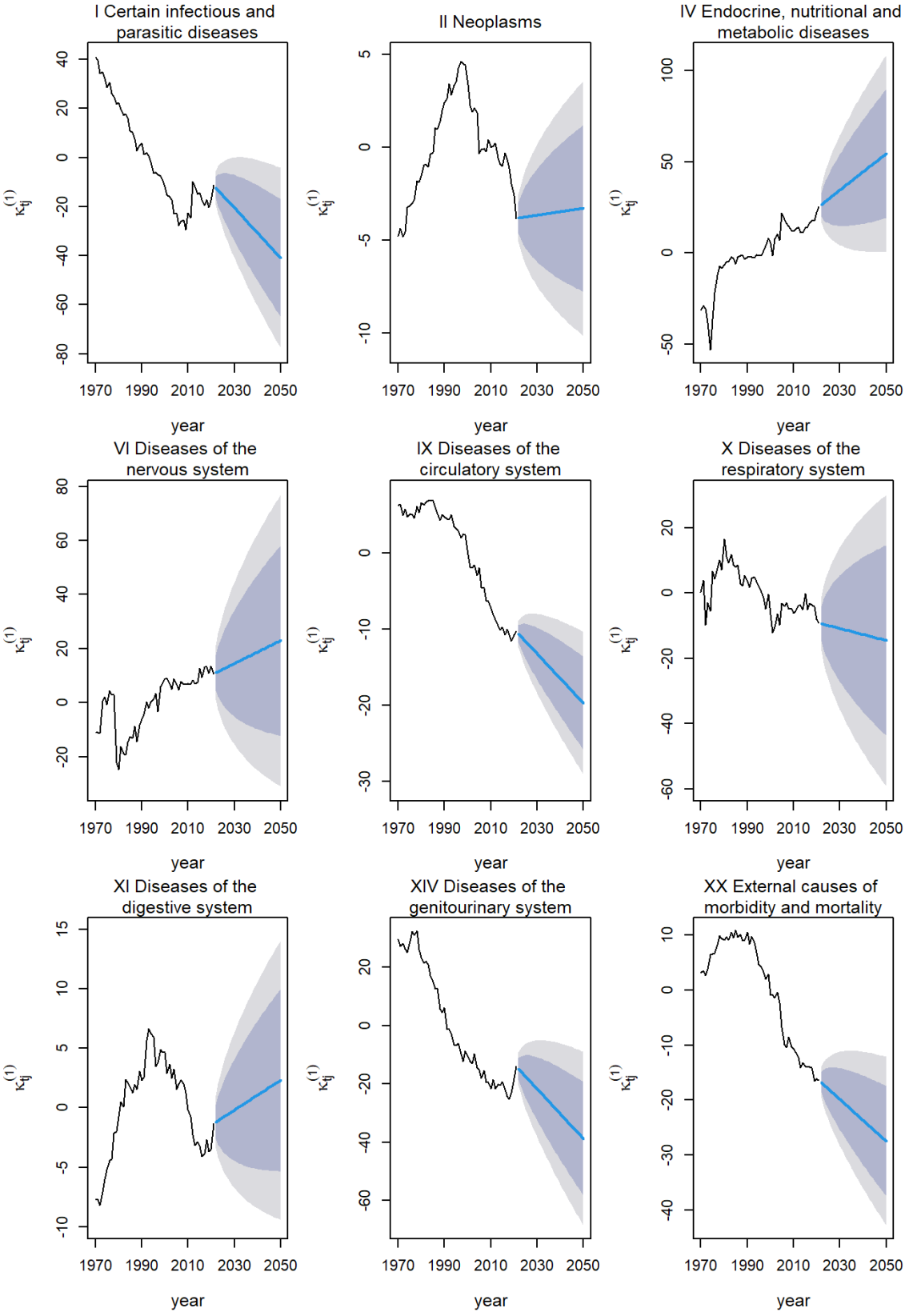
time series can be found in the online Appendix with the parameters $d = 1$ and $p, q = 0$ or 1. Figures 27–28 show the final prediction of the Poisson Lee–Carter model for mortality indices $\kappa_{t(j)}^{(1)}$ by cause of death and sex. The other projections for the other models are presented in the online Appendix.

Table 15. ARIMA models of the mortality indices

The mortality models	$\boldsymbol{\kappa_{t(j)}^{(1)}}$	$\boldsymbol{\kappa_{t(j)}^{(2)}}$	$\boldsymbol{\kappa_{tj}^{(3)}}$
Poisson Lee–Carter/Carter–Lee	ARIMA(0,1,0) with drift	–	–
Binomial Lee–Carter/Carter–Lee	ARIMA(0,1,0) with drift	–	–
Poisson Lee–Carter with two factors (single- and multi-population models)	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift	–
Binomial Lee–Carter with two factors (single- and multi-population models)	ARIMA(0,1,0) with drift	ARIMA(0,1,1) with drift	–
CBD (single- and multi-population models)	ARIMA(0,1,0) with drift	ARIMA(0,1,1) with drift	–
reduced Plat with three factors (single- and multi-population models)	ARIMA(0,1,0) with drift	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
reduced Plat with two factors (single- and multi-population models)	ARIMA(0,1,0) with drift	ARIMA(0,1,1) with drift	–

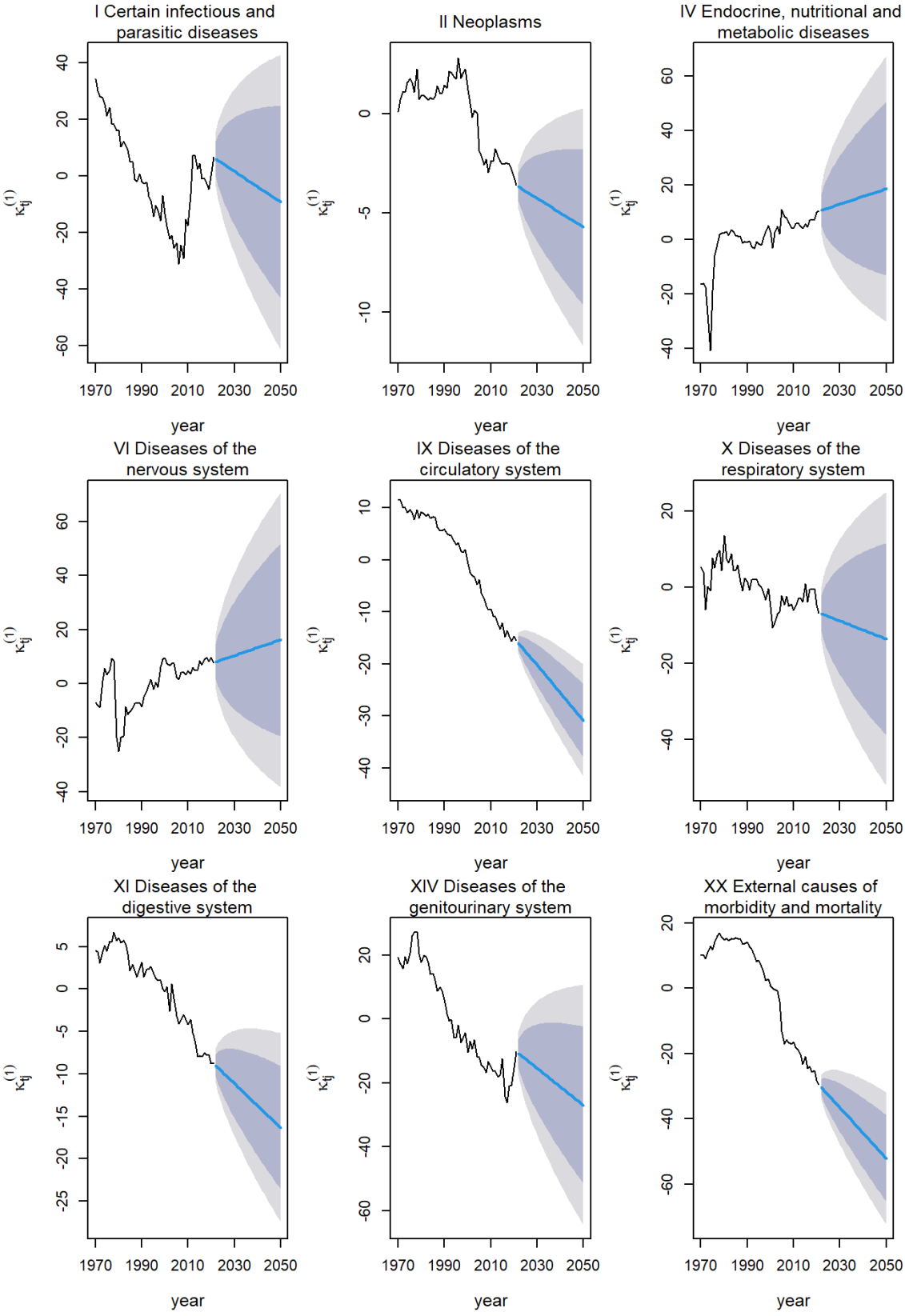
Note: ARIMA models do not differ according to cause of death or sex.
Source: own editing

Figure 27. The group-specific mortality indices of the Poisson Lee–Carter model for males



Note: The gray backgrounds represent the 80% and 95% prediction intervals.
Source: own calculation

Figure 28. The group-specific mortality indices of the Poisson Lee–Carter model for females



Note: The gray backgrounds represent the 80% and 95% prediction intervals.
Source: own calculation

Out-of-sample analysis

Out-of-sample analysis was also used to compare the different mortality models with each other. In order to compare the actual mortality rates with the predicted values, the observation period had to be shortened. This enabled us to create a forecast for a period for which actual data is available. The closer the predicted values are to the actual mortality rates, the more accurate the model. We used the MAPE value to examine this. All mortality models were refitted for the period 1970–2015. Projections were made for the years 2016–2019.[37] All the refitted, age- and time-dependent parameters were similar to those estimated based on the full base period. The ARIMA models that were previously selected were used to forecast the mortality indices. The forecast assumed that the error terms were zero.

Tables 16–17 present the findings of the out-of-sample analysis. According to the out-of-sample analysis, the reduced Plat model with three factors was the most accurate. It predicted mortality rates that were the closest to the current values. From this perspective, the multi-population extension of the reduced Plat model with three factors performed well for neoplasms for males and for diseases of the circulatory system for both sexes. Although the Binomial version of the two-factor Lee–Carter model had previously been shown to be more accurate in an in-sample analysis, the Poisson version of the model produced more accurate predictions.

[37] The years 2020 and 2021 have been omitted because of the COVID-19 pandemic.

Table 16. MAPE values of the mortality models for males, 2016–2019

The mortality models	I	II	IV	VI	IX	X	XI	XIV	XX
Poisson Lee–Carter	0.400	0.082	0.158	0.207	0.047	0.172	0.274	0.344	**0.118**
Poisson Carter–Lee	*0.368*	0.081	–	–	*0.045*	–	0.295	0.358	0.199
Binomial Lee–Carter	0.401	0.087	0.161	0.213	0.056	0.171	0.286	0.347	*0.122*
Binomial Carter–Lee	**0.366**	0.085	–	–	0.052	–	0.308	0.359	0.187
Poisson Lee–Carter with two factors	0.387	0.049	0.149	0.251	0.062	*0.138*	0.135	**0.302**	0.123
Poisson Lee–Carter with two factors (multi-population extension)	0.412	0.051	–	–	0.061	–	0.128	0.324	**0.118**
Binomial Lee–Carter with two factors	0.387	0.057	0.159	0.210	0.064	**0.137**	0.145	*0.311*	0.124
Binomial Lee–Carter with two factors (multi-population extension)	0.428	0.058	–	–	0.063	–	0.137	0.342	*0.122*
CBD	0.435	0.071	**0.145**	0.194	0.105	0.160	0.142	0.325	0.169
CBD (multi-population extension)	n. a.	0.175	–	–	0.136	–	n. a.	n. a.	–
reduced Plat with three factors	0.415	**0.046**	*0.148*	**0.181**	0.095	0.154	*0.125*	0.324	0.134
reduced Plat with three factors (multi-population extension)	0.433	*0.047*	–	–	**0.042**	–	0.152	0.318	0.134
reduced Plat with two factors	0.442	0.050	0.149	*0.183*	0.077	0.164	**0.124**	0.332	0.123
reduced Plat with two factors (multi-population extension)	0.403	0.055	–	–	0.076	–	0.212	0.369	0.194

Notes: The best values are highlighted in bold, and the second-best values are in italics. The abbreviation "n.a." refers to invaluable values, i.e., values that are not available, and indicates a poor fit. The sign "–" denotes that there is no model for the cause of death that has been studied.

Source: own calculation

Table 17. MAPE values of the mortality models for females, 2016–2019

The mortality models	I	II	IV	VI	IX	X	XI	XIV	XX
Poisson Lee–Carter	0.435	*0.053*	0.350	0.171	0.059	0.225	0.156	0.725	0.156
Poisson Carter–Lee	–	–	–	–	0.060	–	0.251	0.855	0.581
Binomial Lee–Carter	0.436	0.062	0.346	0.176	0.061	0.226	0.170	0.722	0.158
Binomial Carter–Lee	–	–	–	–	0.059	–	0.279	0.811	0.552
Poisson Lee–Carter with two factors	**0.319**	0.058	*0.134*	0.165	0.062	*0.118*	*0.117*	**0.523**	0.150
Poisson Lee–Carter with two factors (multi-population extension)	–	–	**0.133**	–	*0.053*	–	0.118	0.608	0.151
Binomial Lee–Carter with two factors	0.462	0.072	0.137	0.183	0.058	0.119	0.125	**0.523**	0.151
Binomial Lee–Carter with two factors (multi-population extension)	–	–	**0.133**	–	0.060	–	0.123	0.586	0.154
CBD	0.448	0.062	0.142	**0.151**	0.125	0.172	0.150	0.597	0.204
CBD (multi-population extension)	–	–	–	–	0.204	–	n. a.	n. a.	n. a.
reduced Plat with three factors	*0.404*	0.054	0.149	0.175	0.055	**0.114**	**0.113**	*0.543*	**0.147**
reduced Plat with three factors (multi-population extension)	–	–	–	–	**0.051**	–	0.141	0.619	0.222
reduced Plat with two factors	0.445	**0.043**	0.159	*0.153*	0.107	0.140	0.125	0.598	*0.148*
reduced Plat with two factors (multi-population extension)	–	–	–	–	0.107	–	0.184	0.771	0.529

Notes: The best values are highlighted in bold, and the second-best values are in italics. The abbreviation "n.a." refers to invaluable values, i.e., values that are not available, and indicates a poor fit. The sign "–" denotes that there is no model for the cause of death that has been studied.

Source: own calculation

The future age-standardized mortality rates by cause of death

Different mortality models can be considered the best for predicting cause-specific mortality based on in-sample and out-of-sample analyses. For this reason, we used all suitable models to forecast the mortality rate of people aged between 60 and 90, and then calculated the ASMR values for the period between 2022 and 2050 using these forecasts. The predicted initial mortality rates can be converted into central mortality rates using Equation 29. This enables the mortality rates predicted by the Binomial and Poisson models to be compared. The future ASMR values were calculated based on central mortality rates.

Figures 29–32 show the ASMR values by cause of death and sex between 1970 and 2050. We refrained from presenting the forecast results of certain models because we did not consider them to be realistic. This was because they did not correspond to previous trends (e.g., the Binomial Lee–Carter model with two factors for II, VI, X, XI, and XX for males and for I, II, VI, X, XI, and XIV for females, and the Binomial Carter–Lee model for XI for males), or prove to be appropriate based on the goodness-of-fit measures. The online Appendix contains further figures and tables of predicted ASMR values.

There are significant differences among the remaining model results. For example, a level shift is evident in the forecast of the Poisson Lee–Carter model with two factors, as well as in the multi-population extension of the reduced Plat model with two factors for certain causes of death among males. The same can be observed for females in both the Binomial Lee–Carter model and the multi-population extension of the reduced Plat model with two factors. The results of the multi-population CBD model were only accepted for diseases of the circulatory system for both sexes. However, even in this case, the results are questionable as they predict a shift in levels, i.e., excess mortality. Based on measures of goodness of fit, out-of-sample analysis and the results of long-term projections, the reduced Plat model with three factors is considered the most suitable for forecasting. It is also one of the best multi-population variants.

Until the mid-1990s, mortality rates for men with regard to neoplasms and digestive diseases deteriorated. The ASMR values calculated for people aged 60–90 were increasing, but this was followed by a significant improvement in mortality rates. However, diseases of the digestive system have stagnated in recent years. Forecasts are particularly difficult in the case of these two causes of death, as it is unclear whether the improvement for men will continue at the same rate in the long term as has been seen

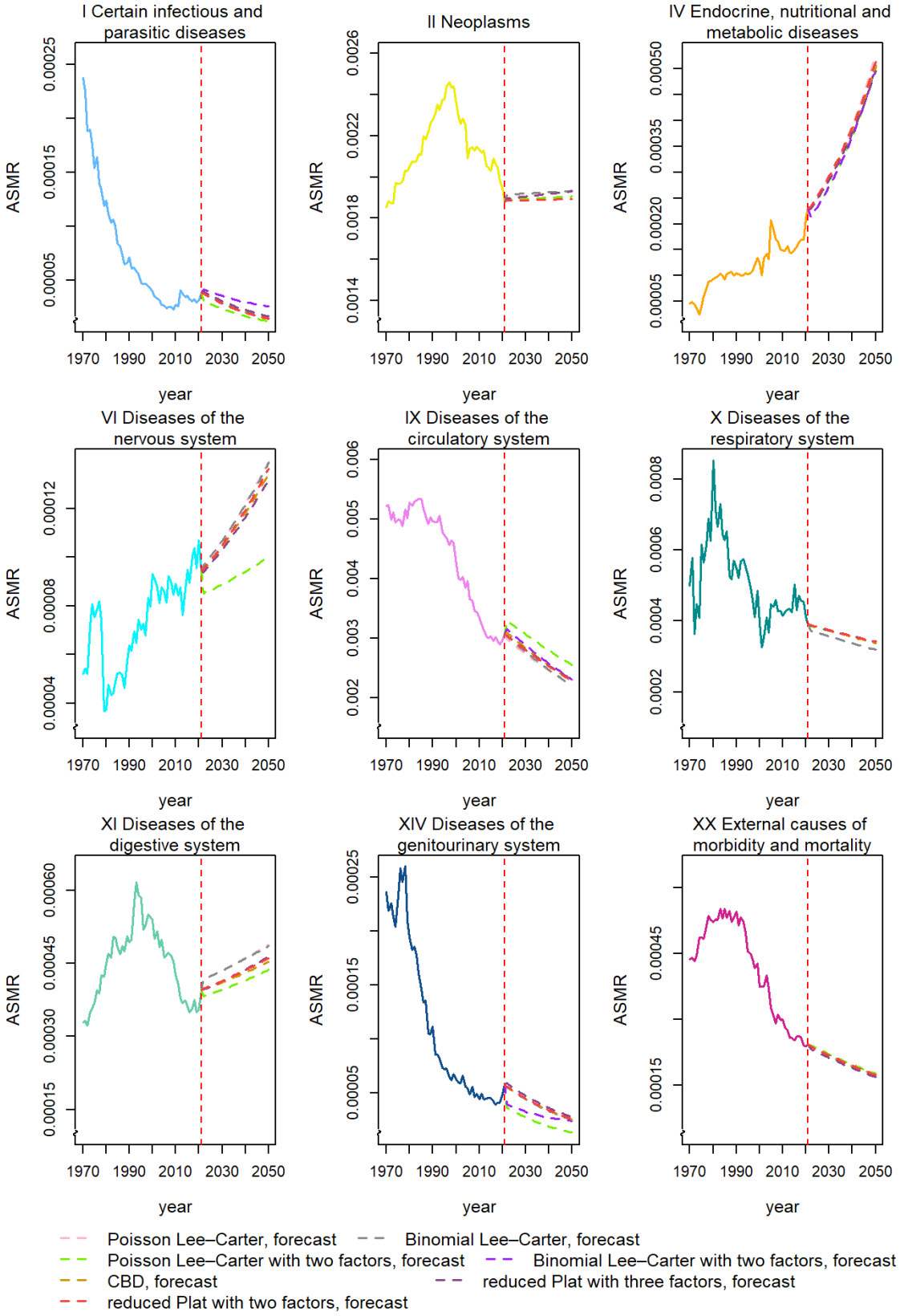
since the second half of the 1990s. It is not necessarily realistic to predict this rate of improvement, as it could lead to an even lower mortality rate in the future than was experienced in the early 1970s. Regarding the base period, ASMR values for women show more predictable trends by cause of death than those for men. However, large fluctuations were observed in the case of endocrine, nutritional and metabolic diseases, as well as diseases of the nervous and respiratory systems. These fluctuations were similar for men.

According to various mortality models, improvements in mortality will not continue for neoplasms and diseases of the digestive system for males. Both sexes are expected to experience a significant worsening of endocrine, nutritional and metabolic diseases, as well as diseases of the nervous system, in the future. For females, there will be an improvement in all other causes of death. The rate of deterioration for endocrine, nutritional and metabolic diseases may accelerate in the future, particularly among men.

Based on the central ASMR values of the base period, it can be seen that men experience higher mortality rates than women for most of the main causes of death (see Figure 5). There is an exception: for decades, mortality rates for women were higher for endocrine, nutritional and metabolic diseases. However, this changed during the second half of the 2010s. Excess mortality was also observed for females for causes I, VI and IX in certain years of the base period. For cause I, this occurred in the 2010s, whereas for causes VI and IX, the examples date back to the 1970s. Since 2005, the mortality gap between men and women has widened for external causes of morbidity and mortality. A similar trend has also been observed for cause IX, but only in recent years.

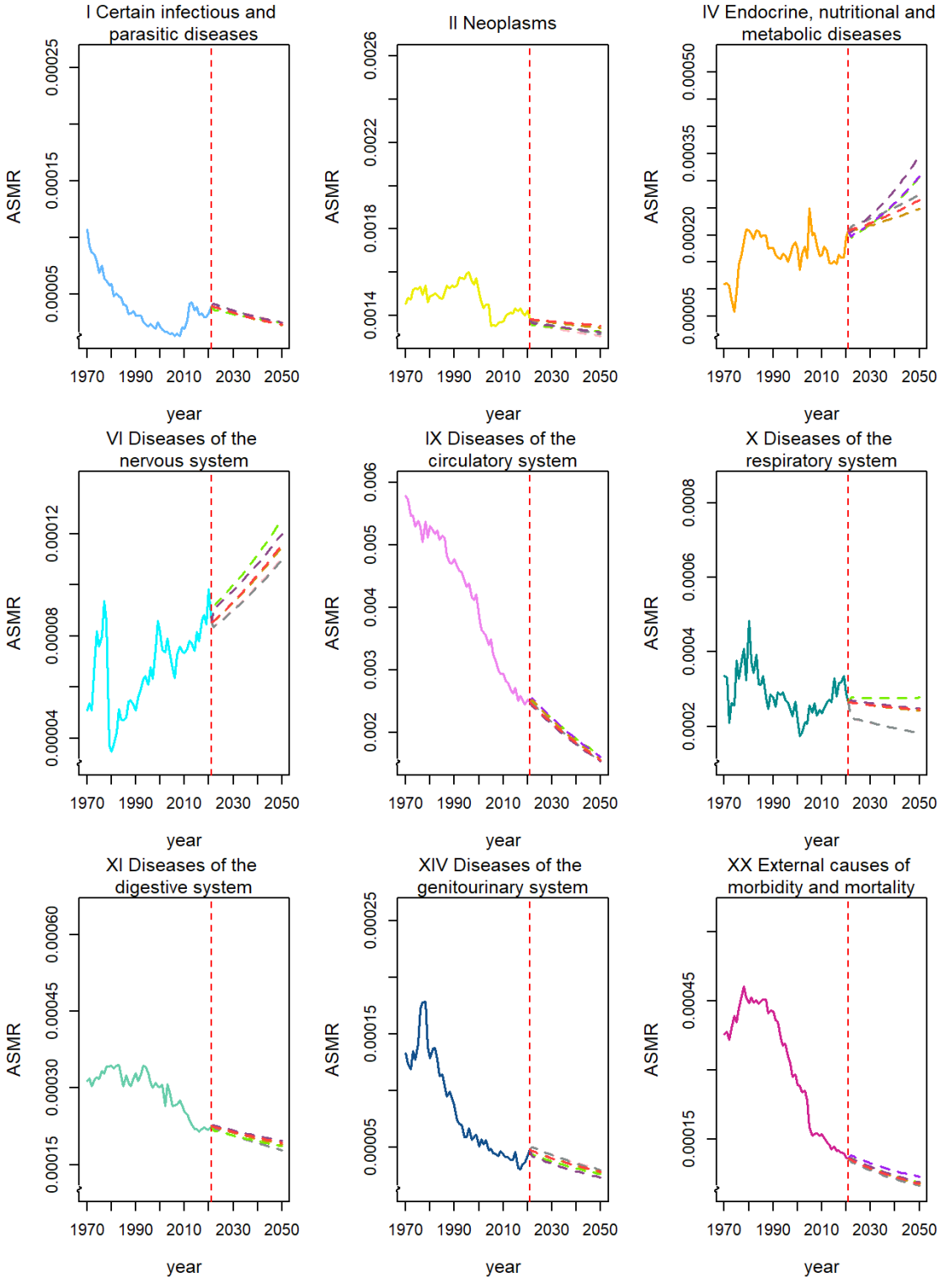
Based on the long-term projection, female mortality is expected to be lower than male mortality for eight out of nine causes of death. According to the reduced Plat model with three factors, women are expected to have a higher mortality rate than men for certain infectious and parasitic diseases. The mortality gap between men and women continues to narrow for cause X, and is narrowing slightly for cause XX. The difference will increase for causes II, IV, IX and XI, but will stagnate for causes VI and XIV. It is possible to compare male and female mortality rates for causes of death IX, XI, XIV and XX using the multi-population version of the reduced Plat model with three factors. This model variant corroborates previous findings.

Figure 29. Age-standardized mortality rates for males between 1970 and 2050 (forecasts are based on single-population mortality models)



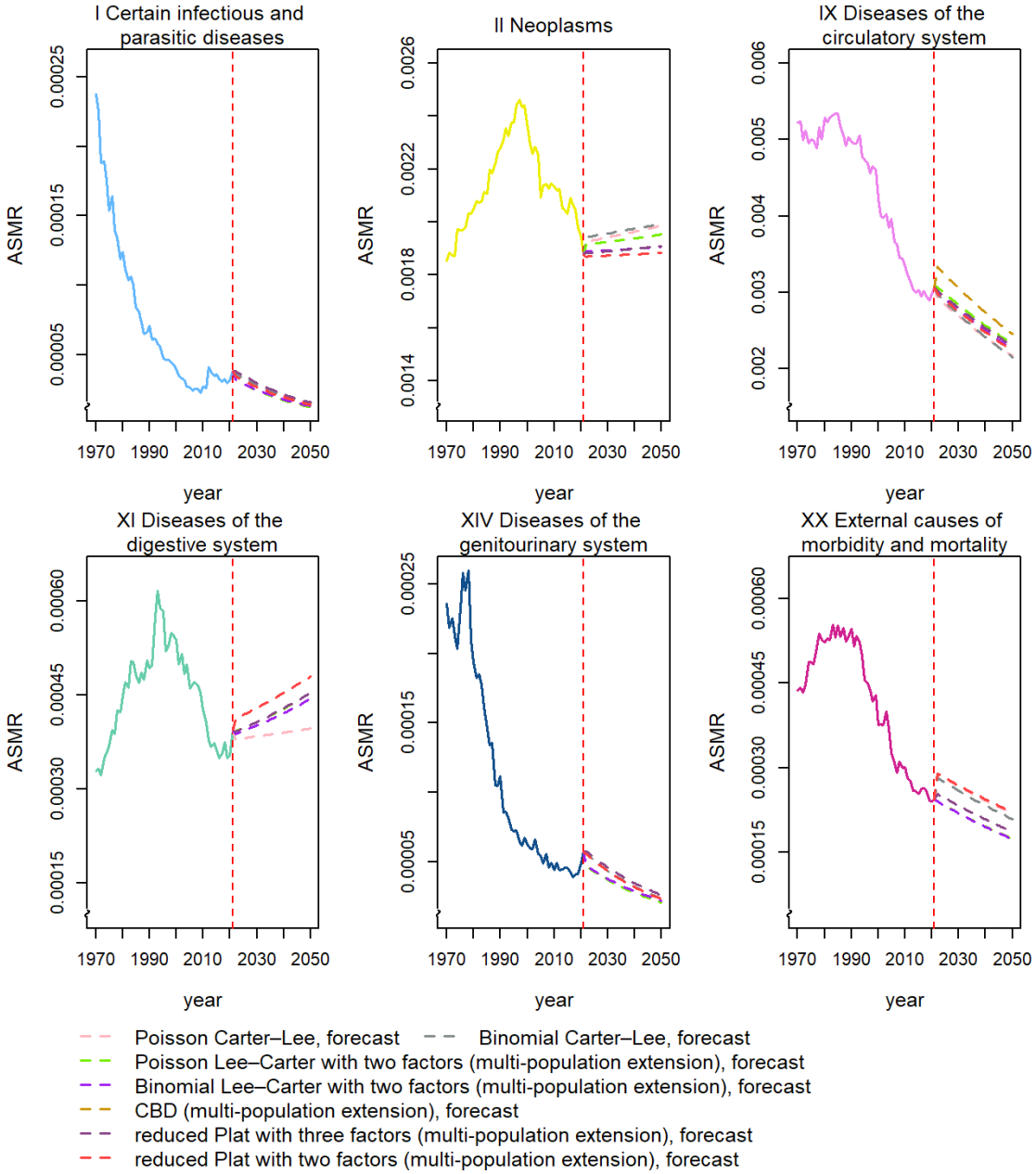
Source: own calculation

Figure 30. Age-standardized mortality rates for females between 1970 and 2050 (forecasts are based on single-population mortality models)



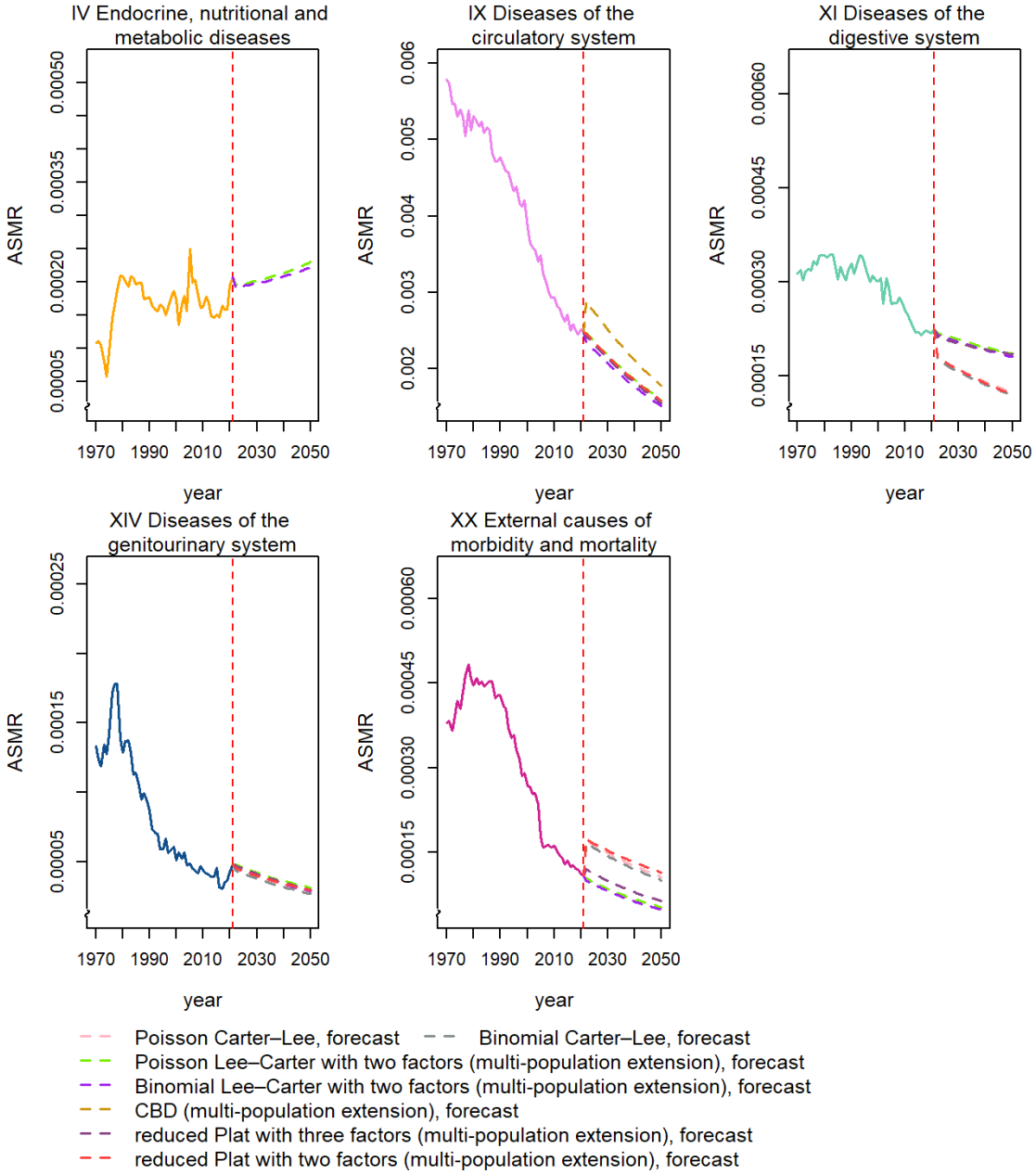
Source: own calculation

Figure 31. Age-standardized mortality rates for males between 1970 and 2050 (forecasts are based on multi-population mortality models)



Source: own calculation

Figure 32. Age-standardized mortality rates for females between 1970 and 2050 (forecasts are based on multi-population mortality models)



Source: own calculation

Summary

The focus of this study was the mortality of Hungarian men and women aged between 60 and 90 over the period from 1970 to 2021. Different stochastic mortality models were fitted according to age, sex, and primary cause of death using the maximum likelihood estimation method with the Newton–Raphson iteration technique. Converted data was used because the methodology for coding causes of death, the ICD classification, has changed several times in recent decades. First, seven single-population mortality models were fitted differing in the following respects: the assumption about the distribution of deaths, the number of mortality indices in the initial equation, and whether the age-varying coefficient(s) were parametric or nonparametric.

The multi-population model variants were developed by clustering the cause-specific mortality indices of the fitted single-population models. Two clustering methods were applied: the hierarchical method and the k-medoids method. The distance between the time series was measured using the dynamic time warping algorithm. The mortality models include those with a large number of parameters and those with a small number of parameters. The single- and multi-population versions of the Poisson/Binomial Lee–Carter model with two factors are the least parsimonious, while the CBD and its multi-population variant are the most parsimonious. The reduced Plat model with three factors and its multi-population variant include the greatest number of stochastic factors in the initial equation. This means that forecasting requires the greatest amount of work in this model. In fact, even two single-population models exhibit common features among subpopulations due to parametric age effects. These are the CBD and the reduced Plat models.

The first mortality index was successfully clustered for the single-population models. The Poisson and Binomial Lee–Carter models with two factors were an exception, since the second mortality index was found to be clustered. Three groups of causes of death showed distinct mortality trends in men: the endocrine, nutritional and metabolic diseases, the diseases of the nervous system, and the diseases of the respiratory system. The other causes of death were successfully clustered for males. The following causes of death formed a single cluster for women: certain infectious and parasitic diseases, neoplasms, diseases of the nervous system, and diseases of the respiratory system.

Multi-population models were used to forecast the mortality of subpopulations. A coherent forecast avoids crossover effects between the age-specific mortality rates of different subgroups. This was not an expectation with regard to multi-population models in this study. The order of the main causes of death may have changed. The number of fitted parameters that need to be predicted can be reduced by using multi-population models. Another advantage is that multi-population mortality models enable the joint analysis of causes of death that are not independent. The crude mortality rates were analyzed. We did not take into account whether the underlying risk factors were the same when clustering causes of death. Different causes of death were grouped into clusters based on the similarity of mortality trends, for which a common trend was then predicted.

The causes of death were categorized differently for males and females. This may be because the lifestyles of men and women were much more different decades ago, for example in terms of alcohol consumption and smoking habits. Therefore, mortality trends for the main causes of death evolved differently. One disadvantage is that changes in diagnosis and coding may affect the evolution of the mortality index(es). Another disadvantage of using multi-population models is that subgroups with a higher number of deaths can influence the mortality index trend more significantly, which affects the forecast.

We prepared a forecast for both sexes up to 2050 according to each model and main cause of death. We predicted all cause-specific and common mortality indices using ARIMA models. Both in-sample and out-of-sample analyses were conducted to compare the mortality models. The Binomial Lee–Carter models with one or two factors and their multi-population extensions were found to be the most suitable models based on the goodness-of-fit measures. However, the fitted parameters of the Binomial models showed smaller or larger differences compared to the Poisson models for certain causes of death. Therefore, their results should be treated with caution when making forecasts, as they may not be reliable. Based on the results of the long-term projection, the Binomial Lee–Carter model with two factors proved to be unrealistic for several causes of death. According to the out-of-sample analysis, the reduced Plat model with three factors appeared to be the most accurate, and its multi-population version was also appropriate for certain causes of death. Overall, this mortality model was well-performing. However, it was not the most appropriate for all causes of death.

Long-term forecasts of the reduced Plat model with three factors showed that the difference in mortality rates between men and women would increase for causes I, II, IV,

IX and XI. Furthermore, according to this model, females will have higher mortality rates from infectious and parasitic diseases than males. For both sexes, mortality will improve to a greater extent for endocrine, nutritional and metabolic diseases, as well as for diseases of the nervous system, based on different models. The improvements in mortality trends observed in the last 25 years may not continue in the future for neoplasms and diseases of the digestive system in the case of men.

IV.3. Rotation of the age-varying parameters in multi-population mortality models[38]

Introduction

In the present study, two multi-population models were selected in addition to the Lee–Carter model[39] to investigate the time-dependence of the age-varying parameters, and to take this into account in the forecast. This is referred to as the rotation of age-dependent parameters. We fitted mortality models to data on Hungarian men and women, analyzing and predicting mortality in these subpopulations. In this study, two augmented multi-population models, the two-factor ACF model and the three-factor ACF model, were fitted using data from 1960 to 2022 and the maximum likelihood estimation method. The results of these models were compared with each other and with the Lee–Carter model. We also constructed rotated versions of all three models. We analyzed mortality rates within the 0–90 age group, treating those aged 90 and over as a single group. Life expectancy at birth for Hungarian men and women is projected to 2050 using rotated and nonrotated models.

Many modified versions of the original log-bilinear Lee–Carter model exist today. Examples include versions with additional factors or multi-population models (see, e.g., Villegas et al. 2017). An innovation is when new methods are used to fit the model or to predict stochastic factors. These types of improvements aim to improve the fit of mortality models, and make the prediction more accurate. For this research topic, we used two versions of augmented models to analyze male and female mortality together, despite the large differences between the two subpopulations. In contrast to the Lee–Carter model, the augmented models allow group-specific factors to be taken into account in addition to the common characteristics of male and female mortality. In these models, the common

[38] Varga (2024) is the basis for this chapter.
[39] In this chapter, Lee–Carter model refers to the Poisson version.

characteristics of the population are described by common time- and age-dependent parameters, while the subpopulation-specific characteristics are described by group-specific parameters. Hungarian men and women are closely related, and have similar socioeconomic backgrounds. The two sexes are exposed to the same influences, for example, they use the same health care system, and nowadays the lifestyles of men and women are less different. This justifies the use of multi-population models, which are often used to predict mortality for men and women together (see, e.g., Carter–Lee 1992, Li–Lee 2005).

The popularity of the Lee–Carter model is supported by the fact that examples of its application can be found in Hungarian research. Baran et al. (2007) fitted the original Lee–Carter model and its two- and three-factor versions using the SVD method to data on Hungarian men and women. The authors have also tried to find the ideal length of the base period to fit the models, as the political and economic regime change in Hungary in 1989–1990 had a significant impact on mortality.[40] Májer–Kovács (2011) applied the Lee–Carter model to predict life expectancy in old age, and to measure the associated uncertainty using Hungarian data. The authors also discussed the risk to the financial and pension system. As well as predicting mortality, the Lee–Carter model can also be used to predict fertility rates (see, e.g., Bajkó et al. 2015, Simpach–Arltova 2016). Bajkó et al. (2015) constructed a pension model in which both mortality and fertility were predicted using the Lee–Carter model. The authors used base period data of different lengths. It was found that for Hungarian data it would be most appropriate to consider data for years after 1980. For validation purposes, projections were made using a set of shortened base period data in order to compare the projected values with the current data.

Vékás (2017, 2019) used the notable models of the generalized age–period–cohort model framework (see, e.g., Villegas et al. 2018) and the *StMoMo* R software package on Hungarian data. Petneházi–Gáll (2019) also considered the Lee–Carter model as a benchmark, and compared the results of the machine learning method with this model. The authors applied the long short-term memory neural network method to predict mortality rates using data from different countries. Gogola–Vékás (2020) analyzed Czech data in addition to Hungarian data, and made an actuarial projection specifically for the older age group. Mortality projections are also the concern of demographers, as well as

[40] According to Józan (2008), the negative impact of regime change was compounded by an increase in slowly progressive diseases.

actuaries. Obádovics–Tóth (2021, 2023) used a modified version of the Lee–Carter model by Lee–Miller (2001) to produce long-term population projections.

Tóth (2021a, 2022a, 2022b) used the Lee–Miller mortality prediction model to predict mortality rates for the period affected by the COVID-19 pandemic, and compared the estimated data with current mortality. The author's aim was to estimate the extent of excess mortality in Hungary. He estimated how mortality would have developed if the COVID-19 epidemic had not occurred. Csiszár (2022) analyzed data by cause of death, partly using Hungarian data. In addition to the Lee–Carter model, the author used another method, the P-splines method. The study by Szentkereszti–Vékás (2022), similar to Petneházi–Gáll (2019), also applied the long short-term memory neural network method to Hungarian data. The Lee–Carter model was also examined to compare the results. Similar to Baran et al. (2007), Petneházi–Gáll (2023) applied the single- and multi-factor Lee–Carter model to Hungarian data, also using the most recent data.

Tóth (2021b) fitted the multi-population product-ratio model developed by Hyndman et al. (2013) using data from the Visegrad Group countries, which included Hungarian data. Multi-population models of the Lee–Carter family were fitted to Hungarian regional data by Varga (2023), who found that the ACF and ACF–CAE models fit the male data well. For women, the Lee–Carter model was found to be the best fit. Szentkereszti–Vékás (2024) analyzed mortality rates in Visegrad Group countries and concluded that the best-trained neural network models outperformed standard mortality models. Lee–Carter and CF models were also fitted for comparison.

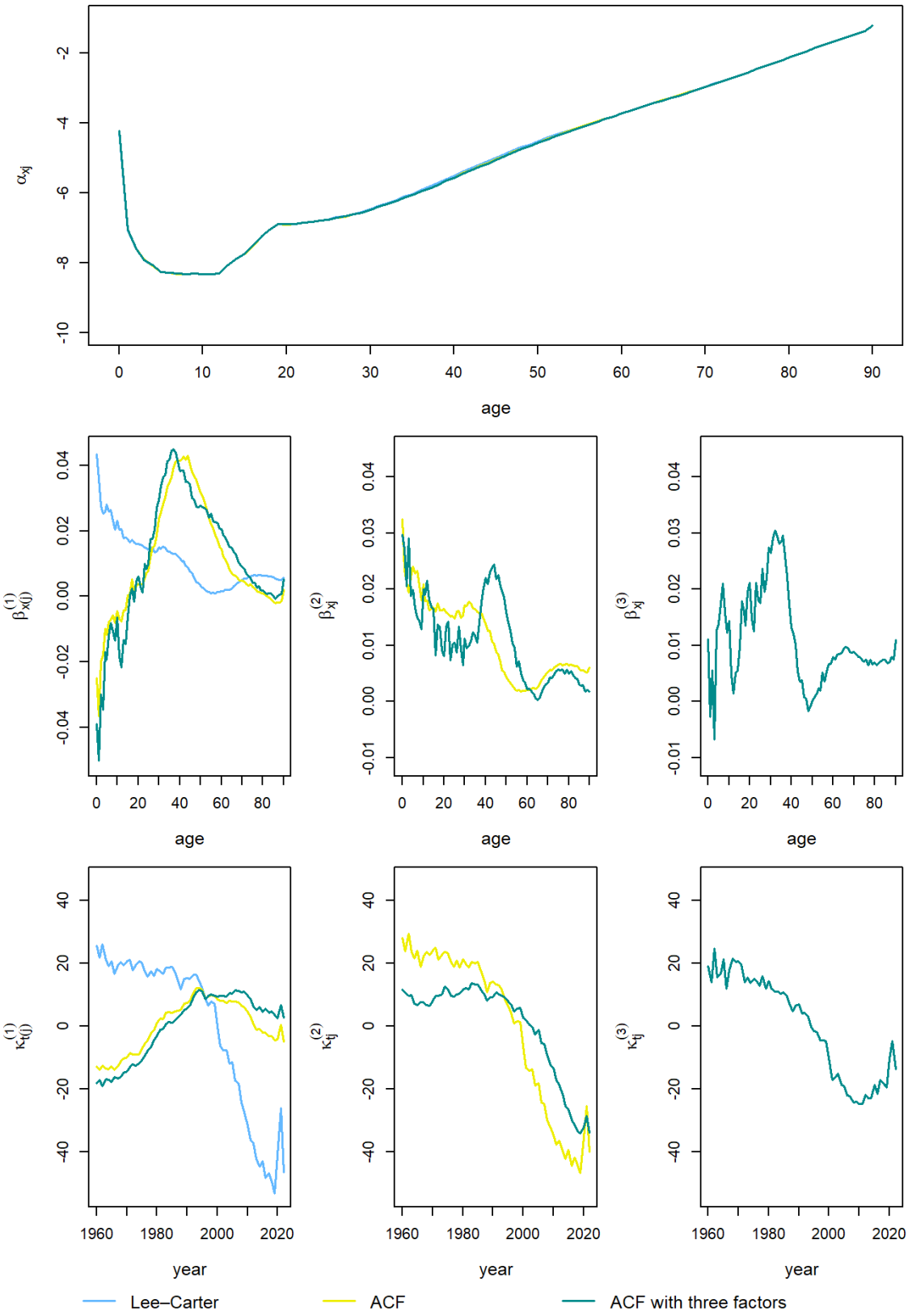
In most of the Hungarian studies mentioned above, the Lee–Carter model was used as a benchmark, and compared with the results of other mortality prediction models. In this study, we chose the ACF model in addition to the Lee–Carter model because it proved to be a good fit for Hungarian regional data (see Chapter IV.1). Examining this model in more detail involves adding an additional factor. Thus, we have created a three-factor version of the ACF model. The novelty of the analysis is that it examines age-dependent parameters as a function of time using a simple approach, which is then incorporated into the forecast. Equations 23–24 were used for rotation, meaning the original initial equations of the different mortality models remained unchanged.

The fitted parameters

Figures 33 and 34 show the fitted parameter values of the three mortality models based on Hungarian data by sex between 1960 and 2022. The ACF and the ACF model with three factors are multi-population variants that include group-specific factors, as well as national, time- and age-dependent parameters. This means that, in addition to modeling common characteristics, they also model sex-specific characteristics (i.e., differences from national characteristics). For these two models, the common parameters are fitted first, followed by the estimation of the nonspecific parameters. The Lee–Carter model captures the mortality characteristics of men and women separately.

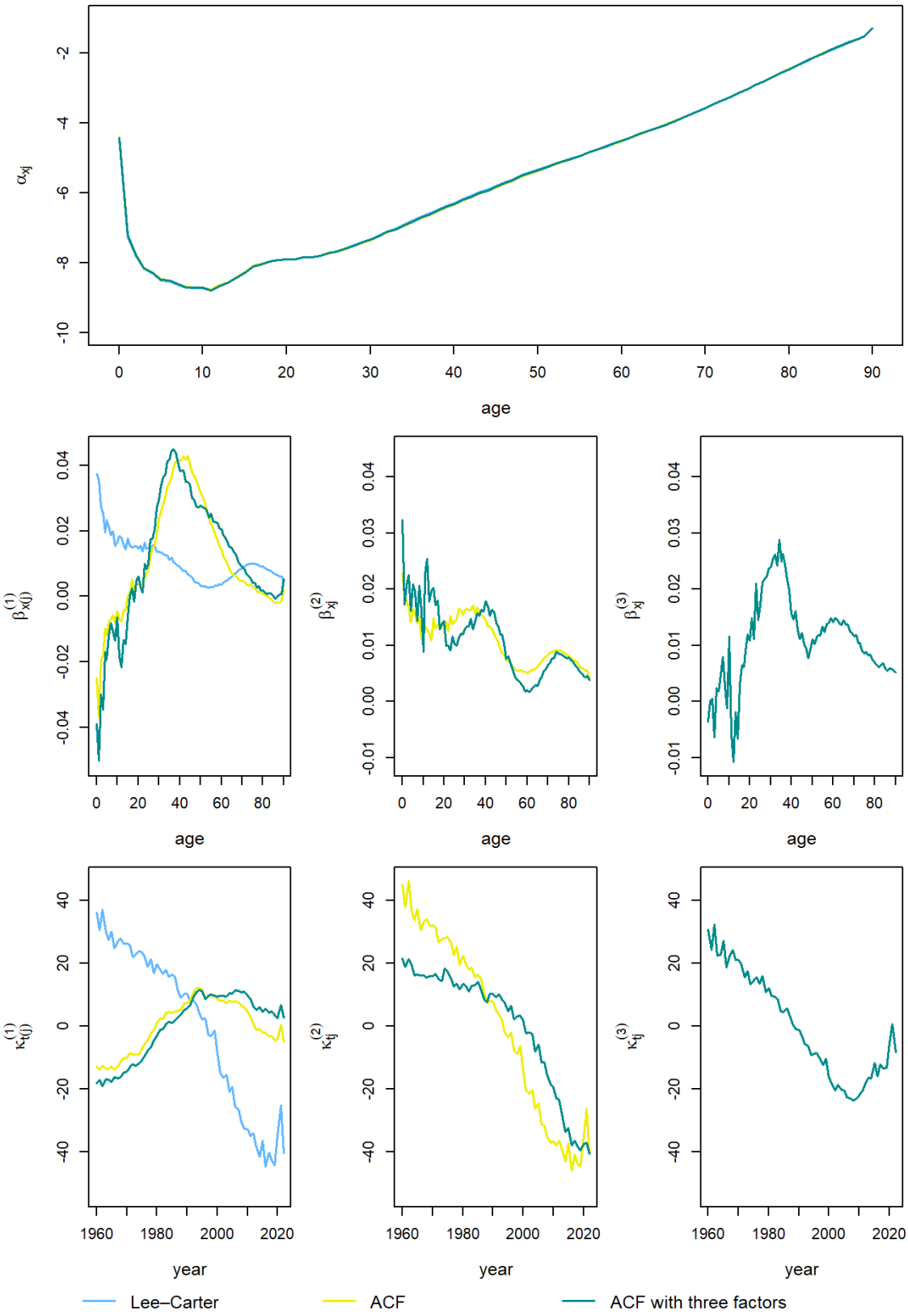
For all three mortality models, the α_{xj} values differ only slightly, and are close to the average of the log mortality rates over the base period. Of the three models, the Lee–Carter model is the most parsimonious, and the ACF model with three factors has the most parameters. The Lee–Carter, ACF and ACF with three factors models allow the trend in mortality change to be described by up to three mortality indices ($\kappa_{t(j)}^{(1)}$, $\kappa_{tj}^{(2)}$, and $\kappa_{tj}^{(3)}$). The greater the number of factors included in a mortality model, the more the trend of the first mortality index will change, as will the shape of the age patterns. Almost all mortality indices show the negative impact of the COVID-19 epidemic on mortality, and the graphs show that mortality improvement slows down in the years before 2020. In the augmented models, the additional indices ($\kappa_{tj}^{(2)}$ and $\kappa_{tj}^{(3)}$) are not as volatile, making them suitable for forecasting. Further figures on the fitted parameters of the mortality models are available in the online Appendix.

Figure 33. The fitted parameters of the mortality models for males



Source: own calculation

Figure 34. The fitted parameters of the mortality models for females



Source: own calculation

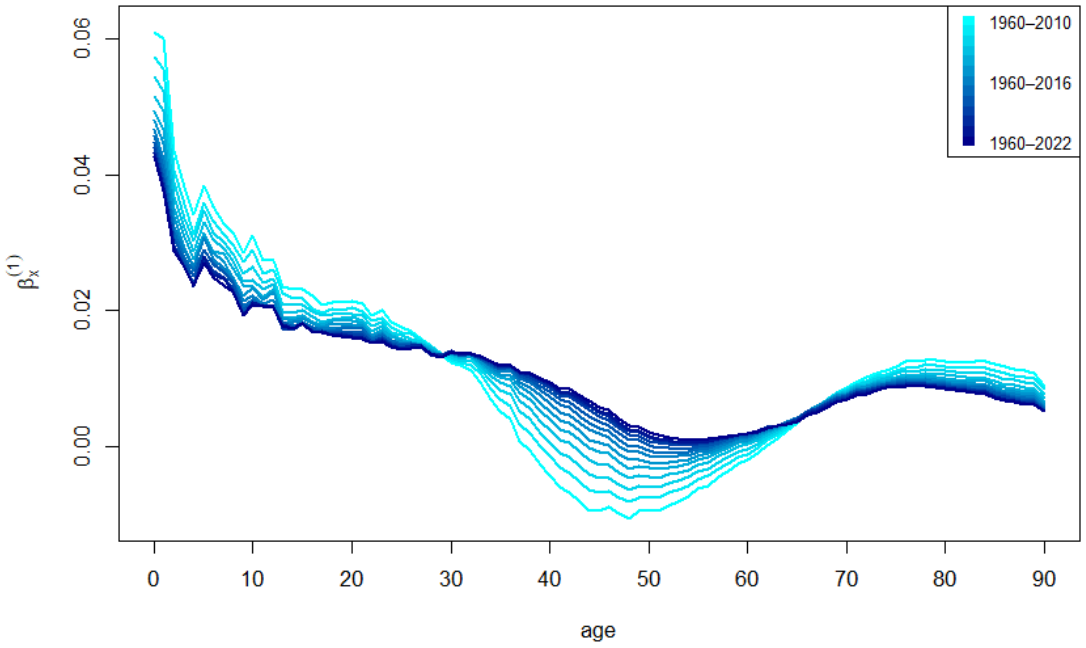
The time-dependence of the age-varying parameters

In order to incorporate the rotation of age-dependent parameters into the predictions of the selected mortality models, as discussed in Chapter I.2.4, it is necessary to analyze how these parameters have changed over time. Equations 23–24 were used to examine this. This requires fitting the models to increasingly shorter base periods, i.e., running the estimation several times. The shortest base period considered in the doctoral research was 1960–2010, and the longest was 1960–2022. The Lee–Carter model for men was the only one in which unstability in the estimate was observed when different base periods of different lengths were used: the model did not converge in all cases. Therefore, it was not possible to generate a rotated version of the Lee–Carter model for men. This was not the case for the augmented models.

Figure 35 shows how the $\beta_x^{(1)}$ parameter changed when base periods of different lengths were used for estimation. These $\beta_x^{(1)}$ parameters are derived from the Lee–Carter model fitted jointly to the male and female data for different base periods. The evolution of sex-specific α_{xj} and $\beta_{x(j)}^{(i)}$ parameters over time is illustrated in other graphs according to different mortality models, which are available in the online Appendix. The graphs clearly show that the rate of change of the age-varying parameter values has slowed down over time.

Figure 35 confirms what we have already discussed in the theoretical section on rotation (see Chapter I.2.4): at the youngest ages, mortality improvement slows down as we see smaller and smaller values of $\beta_x^{(1)}$. The same can be observed for the oldest ages, which is the opposite of what would be expected from the rotation in the forecast. Józan (2008) also hypothesized that the acceleration of mortality improvement in Hungary in the 21st century is most likely to occur in the oldest age group. In comparison, we can see that there is still room for improvement in middle-aged mortality and that this age group can also make a significant contribution to increasing life expectancy in the future. Based on Figure 35, we can conclude that an accelerated improvement in the mortality rate of the oldest age group can still be expected in Hungary.

Figure 35. The shape of the age effect in different time periods for both sexes combined based on the Lee–Carter model



Source: own calculation

The goodness of fit

In the in-sample analysis, we examined the AIC and BIC values for the three mortality models. The AIC and BIC values show that, for both men and women, the ACF model with three factors is the best fitting model, while Lee–Carter is the worst (see Table 18). The mortality models fitted in this study can be derived from each other, i.e., they can be considered as nested models. For this type of model, the likelihood ratio test is more appropriate than information criteria for choosing between models (see, e.g., Plat 2009). Table 19 summarizes the results of the LR tests. The general model fitted better than the nested models, conforming the conclusions drawn from the information criteria.

The magnitude and distribution of the standardized residuals were examined using heatmaps and scatter plots (see the online Appendix).[41] The conclusions were the same as those drawn from the AIC and BIC values or from the LR test: from this point of view, the ACF model with three factors appeared to be the most appropriate of the three models. At the same time, the diagonal pattern in the heatmaps of this model also indicates that the change (improvement) in mortality is not independent of the year of birth. It would

[41] We applied the *StMoMo* package of Villegas et al. (2018) to prepare these figures.

therefore be worthwhile to extend the selected mortality models to include a cohort effect. This is not covered in this study.

Table 18. The mortality models' goodness of fit, 1960–2022

Sexes	Measures	Lee–Carter	ACF	ACF with three factors
Males	AIC	35,773,942	35,695,354	**35,690,129**
	BIC	35,775,559	35,697,982	**35,693,769**
Females	AIC	32,918,361	32,911,775	**32,907,238**
	BIC	32,919,978	32,914,403	**32,910,877**

Note: The best values are highlighted in bold.
Source: own calculation

Table 19. The results of the likelihood ratio tests, 1960–2022

Null hypothesis	Likelihood ratio test statistic		Degrees of freedom	$\chi^2_{J,\alpha}$
	Males	Females		
Lee–Carter model against ACF	78,892	6,892	153	125.4
Lee–Carter model against ACF with three factors	84,420	11,732	307	267.4
ACF against ACF with three factors	5,528	4,840	153	125.4

Note: $\alpha = 0.05$
Source: own calculation

Forecasting the mortality indices

In order to forecast age-specific mortality rates and life expectancy, it is necessary to forecast mortality indices. In this study, ARIMA models were chosen for this prediction. For the two multi-population models, we also assumed independence of the mortality indices within the model and independence of each index between subpopulations. In selecting the appropriate ARIMA(p,d,q) models, we tested for the presence of a unit root in the time series,[42] and checked the autocorrelation and normality of the residuals of the possible ARIMA models.[43] According to the unit root tests, a single differentiation was definitely necessary for all mortality indices. The values 0, 1 and 2 were considered as possible lags for the p and q parameters based on correlograms. In order to choose

[42] We used ADF and KPSS unit root tests.
[43] The following tests were used to check: Ljung–Box Q-statistic, Breusch–Godfrey test, and Shapiro–Wilk test.

between different ARIMA models, we calculated the AIC and BIC values,[44] and examined the width of the prediction intervals. The key statistics for the possible ARIMA models can be found in the online Appendix. Table 20 shows which ARIMA models were chosen. In the projection, the years 2020–2021 were treated as outliers: the effect of the COVID-19 epidemic was additively filtered out. The figures in the online Appendix illustrate the prediction of mortality indices at 80% and 95% prediction intervals.

Table 20. ARIMA models of the mortality indices

The mortality models	$\kappa_{t(j)}^{(1)}$	$\kappa_{tj}^{(2)}$	$\kappa_{tj}^{(3)}$
Lee–Carter	ARIMA(0,1,1) with drift	–	–
ACF	ARIMA(0,1,0) with drift	ARIMA(0,1,1) with drift	–
ACF with three factors	ARIMA(0,1,0) with drift	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift

Note: ARIMA models do not differ according to sex.
Source: own editing

The future life expectancy at birth

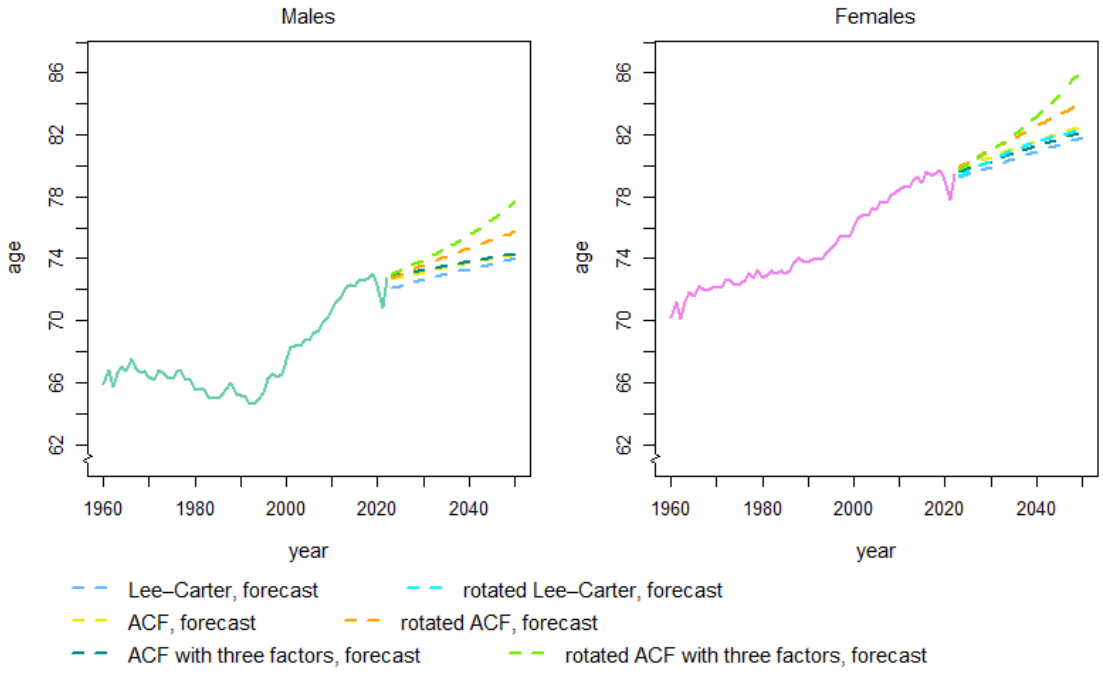
For nonrotated mortality models, prediction of the mortality index(es) is sufficient. For the rotated versions, however, the prediction of age-dependent parameters is also required, as described in Chapter I.2.4 (see Equations 23 and 24). In this study, the rotation was based on the changes observed since 2015, as we found that the rate of change has slowed down over time. Therefore, only the trend in recent years was taken into account for the long-term projection. At the same time, we have assumed that the recent trend will continue in the future. If this is not the case in the long term, our proposed rotation will not be appropriate. This is a weakness of the method used. As the extended models have multiple β_x coefficients, it was important to decide whether rotation was necessary in each case. First, we looked at what happens if all the age patterns in all the models are rotated. This solution proved inadequate, especially in the three-factor case, because life expectancy tended to fall in the long run, contrary to our expectation of a gradual

[44] Expressed by formulas: $AIC = T\ ln(sum\ of\ squared\ residuals) + 2n$ and $BIC = T\ ln(sum\ of\ squared\ residuals) + n\,ln(T)$ where n is the number of estimated parameters ($p + q +$ possible constant term), and T is the number of usable observations (see, e.g., Enders 2014).

improvement in mortality. Finally, we adopted the case where only α_{xj} is rotated in addition to $\beta^{(1)}_{x(j)}$ when predicting mortality models.

The aim was to project life expectancy[45] for men and women up to 2050. Figure 36 shows the results of the projections of the rotated and nonrotated models. Life expectancy at birth is also presented in the online Appendix. The rotated models predicted a higher life expectancy than the nonrotated versions, and this was already reflected in the first year of the forecast. The nonrotated Lee–Carter model predicted the lowest life expectancy for both men and women: 74 and 81.8 years in 2050. Furthermore, according to this model, the longevity gap between men and women will increase. The predictions of the nonrotated ACF model and the nonrotated ACF model with three factors gave almost similar results: 74.3 years for men, 82.6 and 82.2 years for women in 2050. The augmented models also predicted a greater difference in life expectancy between the two sexes than is observed today. Of the rotated models, the ACF model with three factors was the most optimistic. This means that men can expect to live 77.7 years and women 86.3 years at the end of the projection period. At the same time, this model predicted the largest difference in life expectancy between the two sexes.

Figure 36. Life expectancy at birth for males and females between 1960 and 2050



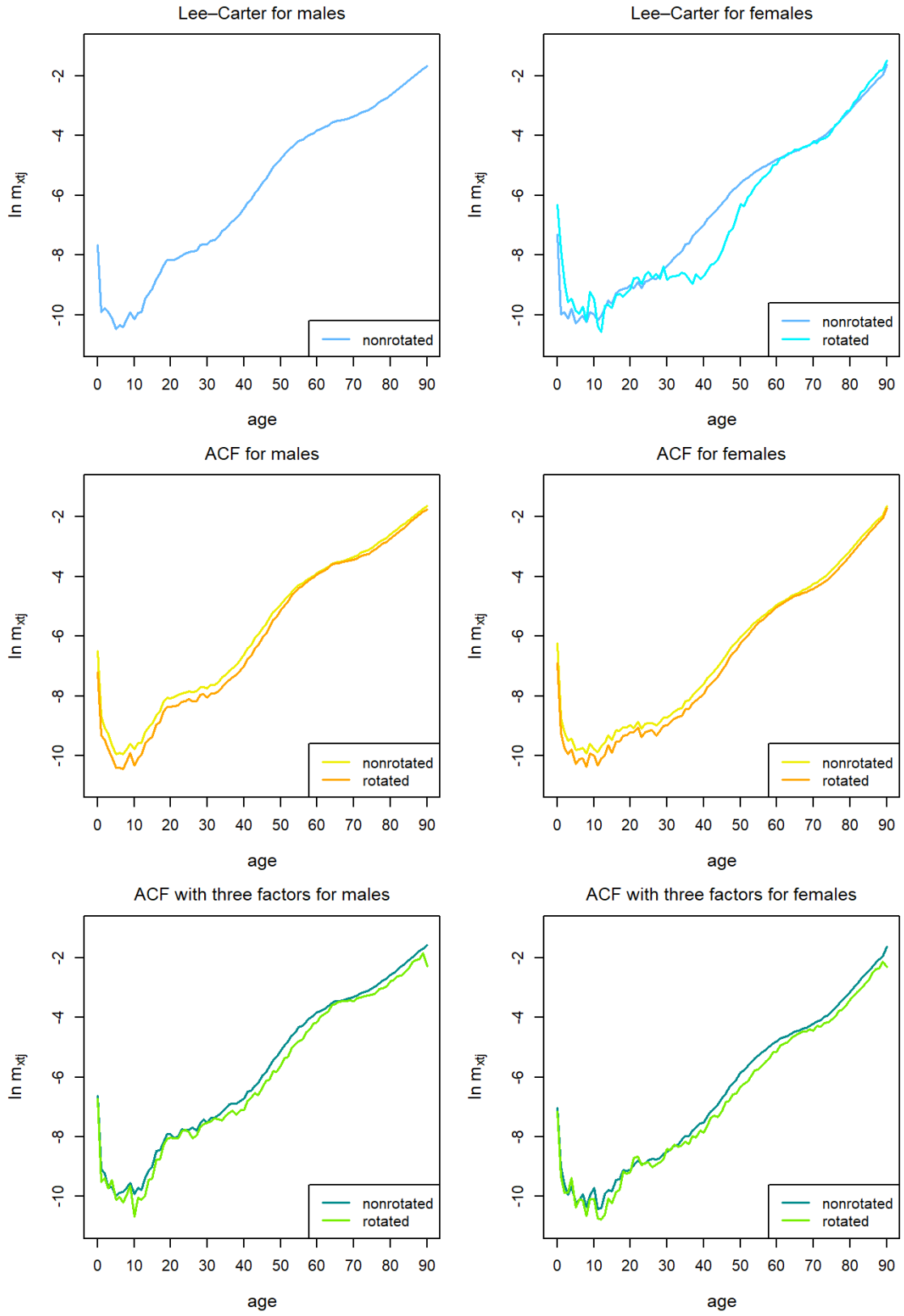
Source: own calculation

[45] In this study, life tables were calculated using the methodology of Chiang (1968) and Eurostat (see https://ec.europa.eu/eurostat/cache/metadata/Annexes/demo_mor_esms_an_1.pdf), except that the group of people aged 90 and over was considered as the oldest cohort instead of people aged 85 and over.

With regard to rotation, an important question is which age groups are likely to contribute more to the increase in life expectancy in the future. Based on Figure 37, our preliminary expectation was that the rotated Lee–Carter model would be expected to improve mortality more in the middle-aged groups. This is more difficult to assess for the augmented models, as they include several age-dependent coefficients. Figure 37 shows how log mortality rates are expected to evolve in 2050 for men and women separately according to the different models. As mentioned above, the rotated Lee–Carter model was only fitted to women. According to this estimate, the rotated Lee–Carter model actually led to improvements in middle-age mortality compared with the nonrotated version, and this led to a significant improvement by 2050. However, for the youngest and oldest age groups, mortality according to the rotated Lee–Carter model was worse than according to the nonrotated version. Thus, the rotated Lee–Carter model seems to be a less suitable prediction model in the long run, based on our rotation methodology. The online Appendix provides the logarithm of the central mortality rates by sex and model for 2050.

The rotated ACF model shows a different result from the rotated Lee–Carter model. Almost all age groups (except those aged 60 and over) contribute to the increase in life expectancy by 2050 according to the rotated ACF model compared to the nonrotated ACF model. The ACF model with three factors corresponds most closely to what was originally expected from rotation. In the youngest age groups, there is no significant difference between the rotated and nonrotated versions, while all other age groups show a more favourable future improvement in mortality, especially those aged 90 and over. As a result, life expectancy is projected to increase at an accelerating rate over the long term to 2050 according to the rotated ACF model with three factors (see Figure 36).

Figure 37. The logarithm of the central mortality rates in 2050 based on different mortality models



Source: own calculation

Coherence analysis

The Lee–Carter model predicted male and female mortality independently. The initial equations of the two augmented multi-population models contain common parameters, which could cause the predicted mortality rates of the groups to converge. Coherent (see Li–Lee 2005) multi-population models imply that mortality differences between subpopulations observed in the past, even if due to biological endowments, will not change in the future. For example, in developed countries such as Hungary, the infant mortality rate for girls is usually lower than that for boys. We do not expect this trend to be reversed in the prediction of mortality. However, by independently predicting the mortality rates of subpopulations, a crossover effect is possible.

By comparing predicted age-specific mortality rates by subpopulation, we can determine whether the difference in mortality between groups will change in the future according to a given model. To examine this, the age-specific mortality rates of the groups can be divided by each other. In this study, we calculated the female-to-male ratio, i.e., the death rates for women divided by the death rates for men. The values of these ratios are available in the online Appendix. The comparison shows that, for all forecast years, the predicted infant mortality rate is higher for women than for men in both the Lee–Carter and ACF models. However, this cannot be assumed. By contrast, the ACF model with three factors performs well in this respect. In addition, the female-to-male ratio for the oldest age group is also appropriate: female mortality does not increase relative to male mortality. This is consistent with what would be expected from the data for the current period. The Lee–Carter model shows less change in the female-to-male ratio over time. For the augmented models there are several age groups where the ratio does not remain constant.

Summary

Three mortality models were fitted in this study: the Lee–Carter model and two multi-population versions, the ACF model and the ACF model with three factors. Data for Hungarian men and women between 1960 and 2022 were used to fit the models. Rotated versions of the three models were also developed for predicting mortality, i.e., the time-dependence of certain age-varying factors was taken into account in the prediction of mortality rates. This change is known in the literature as rotation. In Hungary, it can be

assumed that the improvement in mortality for the youngest age groups is slowing down, i.e., their reserves are being used up, while life expectancy for the oldest age groups is improving. Accounting for rotation in projections may be particularly important as increasing life expectancy, and longevity risk pose challenges to the pension system.

Rotation is typically considered in the context of the Lee–Carter model. In this research, however, we incorporated the time-dependence of age-varying parameters into predictions for two multi-population models. This can be considered a novelty in the literature. Age-specific mortality rates have been projected by sex up to 2050. It is not so simple to compare our results with those found in other studies. This is partly because we used a different method, and partly because few studies have examined rotation using Hungarian data so far. Vékás (2018) was the first to analyze the need for rotation in Hungary using the method developed by Li et al. (2013) (see Chapter I.2.4). Furthermore, according to Vékás (2020), the rotation of the age effect is important in 11 of the 28 European Union Member States for both sexes.

Based on our research results, the Lee–Carter model fitted to Hungarian data does not currently indicate a need for rotation in the short term. Of the nonrotated models, the ACF with three factors model was found to be the best fit, and for this model the infant and old-age mortality rates were found to be appropriate based on the female-to-male ratios. The rotated version of the three-factor ACF model has been successful, i.e., it has fulfilled what we expect from the concept of rotation: in the future, the older age group contributes more to the increase in life expectancy than the youngest. In the long run, however, life expectancy at birth increases at an accelerating rate according to the rotated ACF with three factors model, which is a drawback because it does not seem plausible.

Given the Lee–Carter model, the rotation method used in this research may only be appropriate for short-term forecasting, as it significantly improves mortality in middle-aged groups, which seems unrealistic compared to mortality in other age groups. The present study also found that the Lee–Carter model did not converge on the data for men in each of the shortened base periods, so it was not possible to examine the change in age-dependent parameters over time in their case using the historical data. Rotation was examined not only for the age-dependent coefficients of the mortality index(es), but also for the baseline shape of mortality. Rotation of the latter parameter may be an alternative to the Lee–Miller (2001) method to avoid jump-off bias. It can be argued that it is important to take rotation into account in the projection, as historical data show that different age groups show different rates of mortality improvement. Further research

could be devoted to developing a rotation method suitable for (multi-population) mortality models (with multiple factors) in both the short and long term.

IV.4. Forecasting multi-population mortality models using tree-based machine learning methods

Introduction

The use of machine learning methods in mortality forecasting is a relatively new approach and not as widespread as the standard stochastic models that evolved from the model introduced by Lee–Carter (1992). In this study, we applied tree-based machine learning methods, such as bagging, random forests and gradient boosting machine (GBM), to forecast the long-term mortality of Hungarian men and women. These machine learning methods were used to improve the projections of the three stochastic models, the Lee–Carter,[46] the ACF, and the ACF–CAE models. We fitted the machine learning methods to the residuals of the standard models, examining their dependence on year of birth, and produced a long-term forecast that was incorporated into the forecast of the three standard models. The predictions of these combined models were compared with the results of the standard model versions. The models were fitted using Hungarian mortality data from 1960 to 2022. Mortality rates for people aged 0–90 were analyzed and projected up to the year 2050. This study considers people aged 90 and over as a whole age group. The ACF and ACF–CAE models were selected because they proved to be more suitable than other models when fitted to Hungarian regional data separately by sex (see Chapter IV.1). The Lee–Carter model is the main benchmark model. The two augmented models are extensions of it. In this study, multi-population mortality models were applied jointly to both sexes. This means that men and women represent subpopulations.

Generalized age–period–cohort models typically decompose the long-term time series of mortality rates into two or three dimensions. These are the age, period, and cohort dimensions. This allows the change in mortality to be examined as a function of these factors. In order to forecast the values of future mortality rates, it is necessary to forecast the parameters that describe the trend and the cohort effect. Time series analysis methods are typically applied for this purpose. The Lee–Carter, ACF, and ACF–CAE models do

[46] In this chapter, Lee–Carter model refers to the Poisson version.

not quantify cohort effects. Therefore, our goal in building the combined models was to train the machine learning models to quantify the cohort effect on the residuals of the fitted standard models. For these three conventional models, therefore, additional versions were created in which the cohort effect remaining in the error term was captured using machine learning methods. Two additional variants were created for each of the traditional models: random forests and gradient boosting combined forecasting. Standard mortality models that quantify the cohort effect can be problematic if the time series of the cohort effect is highly volatile, making the prediction uncertain. To avoid this, we applied combined models. The novelty of the analysis is that the results of tree-based machine learning methods have been used for long-term forecasting, extending beyond the end of the observed data series.

One limitation to the widespread adoption of machine learning methods is that they are often black-boxes (especially neural networks) and sensitive models in terms of finding optimal hyperparameters, so overfitting on the training data set can be a problem. The first difficulty mentioned above led us to choose tree-based machine learning methods, which are among the more interpretable methods. In the case of machine learning methods, the explanatory variables can include factors that are not included in standard mortality models (e.g., educational attainment). Another advantage is that these methods can also be used to capture nonlinear relationships. This study applies tree-based machine learning methods, considering only one or two explanatory variables: the cohort and a dummy variable indicating the years in which the COVID-19 epidemic significantly impacted mortality.

In the field of mortality forecasting, a growing body of literature applies machine learning methods to improve the predictive accuracy of standard stochastic mortality models. Deprez et al. (2017) examined two standard mortality models – the Lee–Carter model and its cohort-extended version (see Renshaw–Haberman 2006) – to study how their respective strengths and weaknesses can be revealed using regression tree boosting machine method. Richman (2018) investigated which areas of actuarial science could benefit from the application of artificial intelligence and machine learning methods, mentioning mortality forecasting as an example. Levantesi–Pizzorusso (2019) used decision tree, random forest, and GBM methods to calibrate a machine learning estimator predicted according to the Lee–Carter framework. Their goal was to enhance the quality of forecasts provided by standard models.

Nigri et al. (2019) chose a long short-term memory (LSTM) architecture over ARIMA models for predicting the mortality index of the Lee–Carter model. They found that the LSTM model produced more accurate predictions. Richman–Wüthrich (2019) applied recurrent neural networks (RNNs). They demonstrated how the two most popular RNN architectures – LSTM networks and gated recurrent unit (GRU) networks – worked by using them to forecast mortality rates directly. These architectures are suitable for modeling time series data. In particular, the authors found that the LSTM architecture outperforms the Lee–Carter model for mortality prediction. As the results of the neural network regression models were not sufficiently robust, the use of ensemble methods was recommended. The random forest method is combined with two-dimensional P-spline in Levantesi–Nigri (2020). Their aim was to forecast the random forest estimator of the ratio between the observed and the estimated deaths provided by the Lee–Carter model. The P-spline model was used to smooth and predict the machine learning estimator.

Marino et al. (2021) addressed estimation uncertainty in mortality predictions, filling a gap in the machine learning and deep learning literature. The authors used the LSTM method and estimated the variance of their model using the bootstrap strategy. Perla et al. (2021) primarily focused on the convolutional neural network (CNN) method, but also adapted LSTM networks to analyze mortality data and achieve highly accurate forecasts. Richman–Wüthrich (2021) used deep neural networks to study the mortality of multiple populations. They found that neural networks were more accurate at predicting mortality than standard models.

Our research follows the approach of Bjerre (2022). Alongside various standard stochastic mortality models, Bjerre (2022) used pure random forests and pure GBM methods to forecast log mortality rates. The author also developed stochastic models by combining them with random forests and GBM methods, using the residuals of standard models. The neural networks approach played an important role in forecasting mortality rates in the study by Schnürch–Korn (2022). They used feed-forward neural network (FFNN), RNN, and CNN models. They also compared these with Lee–Carter and ACF models. They found that the CNN model performed best. Wüthrich–Merz (2023) provide an excellent summary that includes a detailed discussion of neural network methods for predicting mortality.

Some research in the field of mortality forecasting uses machine learning methods in the Hungarian literature. Petneházi–Gáll (2019) applied LSTM networks to forecast mortality rates based on data from several countries, including Hungary. They found that

the LSTM method produced more accurate forecasts than the Lee–Carter model. Szentkereszti–Vékás (2022) used LSTM networks to model Hungarian mortality data. The results were compared with the Lee–Carter model forecast. According to their findings, the LSTM model provides significantly more accurate forecasts for the young and middle-aged population. Szentkereszti–Vékás (2024) predicted mortality rates in the Visegrad Group countries using recurrent neural networks. The application of RNN, LSTM and GRU architectures was examined. For comparison, Lee–Carter and CF models were also fitted. According to the researchers, the best-trained neural network models achieved the best performance in forecasting. The following section describes the procedures of the methods we use.

The algorithms of the models

In this study, the stochastic mortality models were fitted using the maximum likelihood estimation method with Newton–Raphson iteration. In the iterative updating scheme, iteration continued until the increase in the log-likelihood function was less than 10^{-6}. The parameters of the models were determined simultaneously, taking into account the necessary constraints. We also created weighted versions of the Lee–Carter, ACF and ACF–CAE models. This means that if the number of deaths in a given year and age group did not reach five, it was not taken into account in our calculations. This is different from the weighted model described by Wilmoth (1993). The author developed the weighted SVD method, which uses the observed number of deaths for each age as a weight. In our research, we used maximum likelihood estimation. We created a weight matrix for fitting the weighted versions of the mortality models. The w_{xtj} weight matrix has dimensions of $x \times t$ where x denotes the age, t is the year, and j refers to the subpopulation. The matrix elements take values of 0 or 1, indicating which cells should be ignored for each age group and year. The log-likelihood functions of the weighted mortality models are reviewed below.

The log-likelihood functions for the weighted standard mortality models can be written as follows:

weighted Lee–Carter

$$\ell_j(\theta) = \sum_{xt}\left\{w_{xtj}\left(D_{xtj}\,(\alpha_{xj} + \beta_{xj}^{(1)}\kappa_{tj}^{(1)}) - E_{xtj}^{c}e^{\alpha_{xj}+\beta_{xj}^{(1)}\kappa_{tj}^{(1)}}\right)\right\} + \mathrm{C} \tag{45}$$

weighted ACF

$$\ell_j(\theta) = \sum_{xt}\left\{w_{xtj}\left(D_{xtj}\,(\alpha_{xj} + \beta_{x}^{(1)}\kappa_{t}^{(1)} + \beta_{xj}^{(2)}\kappa_{tj}^{(2)}) - E_{xtj}^{c}e^{\alpha_{xj}+\beta_{x}^{(1)}\kappa_{t}^{(1)}+\beta_{xj}^{(2)}\kappa_{tj}^{(2)}}\right)\right\} + \mathrm{C} \tag{46}$$

weighted ACF–CAE

$$\ell_j(\theta) = \sum_{xt}\left\{w_{xtj}\left(D_{xtj}\,(\alpha_{xj} + \beta_{x}^{(1)}\kappa_{t}^{(1)} + \beta_{x}^{(2)}\kappa_{tj}^{(2)}) - E_{xtj}^{c}e^{\alpha_{xj}+\beta_{x}^{(1)}\kappa_{t}^{(1)}+\beta_{x}^{(2)}\kappa_{tj}^{(2)}}\right)\right\} + \mathrm{C} \tag{47}$$

where C is a constant.

After fitting the stochastic mortality models, it was possible to create combined models. Random forests and GBM methods were then applied to predict the residuals of the Lee–Carter, ACF and ACF–CAE models. We used two different approaches related to machine learning methods: in one case, the cohort was the only explanatory variable, while in the other case, a dummy variable was incorporated into the model alongside the cohort. The value of the dummy variable indicated whether the pandemic had significantly impacted mortality in a given year, specifically in 2020 and 2021. We repeated the procedure using the residuals from the weighted versions of each stochastic mortality model.

The process for building combined stochastic mortality models is as follows:

1. fitting stochastic mortality models
2. computing residuals: difference between the actual and the fitted log mortality values
   $res_{xtj}^{model} = ln\, m_{xtj} - \widehat{ln\, m}_{xtj}^{model}$
3. applying random forests or GBM to residuals in different training data sets
   $res_{xtj}^{model} \sim cohort$ or $res_{xtj}^{model} \sim cohort + COVID\ dummy$
4. optimization of the hyperparameters using different test data sets
5. forecasting the residuals for H-period ahead, h = 1, …, H
   $\widehat{res}_{xj,t+h}^{model,RF}$
   $\widehat{res}_{xj,t+h}^{model,GB}$
6. forecasting the mortality indices for H-period ahead (h = 1, …, H) in order to predict log mortality rates according to standard stochastic models
   $\widehat{ln\, m}_{xj,t+h}^{model}$
7. forecasting the combined stochastic mortality models
   $\widehat{ln\, m}_{xj,t+h}^{model,RF} = \widehat{ln\, m}_{xj,t+h}^{model} + \widehat{res}_{xj,t+h}^{model,RF}$
   $\widehat{ln\, m}_{xj,t+h}^{model,GB} = \widehat{ln\, m}_{xj,t+h}^{model} + \widehat{res}_{xj,t+h}^{model,GB}$

In summary, there were three standard stochastic models and we fitted weighted versions of all of them. In each case, we created additional models by combining the weighted and unweighted mortality models with two types of tree-based machine learning method. We also fitted the latter versions in two ways, depending on the number of explanatory variables. Thus, we were able to make predictions based on a total of 30 models. The *randomForest* R package[47] was used to implement the random forests method, while the *gbm* R package[48] was used for stochastic GBM. Table 21 summarizes the mortality models that were fitted.

[47] See https://CRAN.R-project.org/package=randomForest.
[48] See https://CRAN.R-project.org/package=gbm.

Table 21. Summary of the fitted standard and combined mortality models

The types of the models	The shortened names of the models
standard stochastic models	unweighted/weighted Lee–Carter
	unweighted/weighted ACF
	unweighted/weighted ACF–CAE
random forests combined with stochastic models	unweighted/weighted Lee–Carter + cohort effect (random forest)
	unweighted/weighted ACF + cohort effect (random forest)
	unweighted/weighted ACF–CAE + cohort effect (random forest)
	unweighted/weighted Lee–Carter + cohort and COVID effects (random forest)
	unweighted/weighted ACF + cohort and COVID effects (random forest)
	unweighted/weighted ACF–CAE + cohort and COVID effects (random forest)
GBM combined with stochastic models	unweighted/weighted Lee–Carter + cohort effect (GBM)
	unweighted/weighted ACF + cohort effect (GBM)
	unweighted/weighted ACF–CAE + cohort effect (GBM)
	unweighted/weighted Lee–Carter + cohort and COVID effects (GBM)
	unweighted/weighted ACF + cohort and COVID effects (GBM)
	unweighted/weighted ACF–CAE + cohort and COVID effects (GBM)

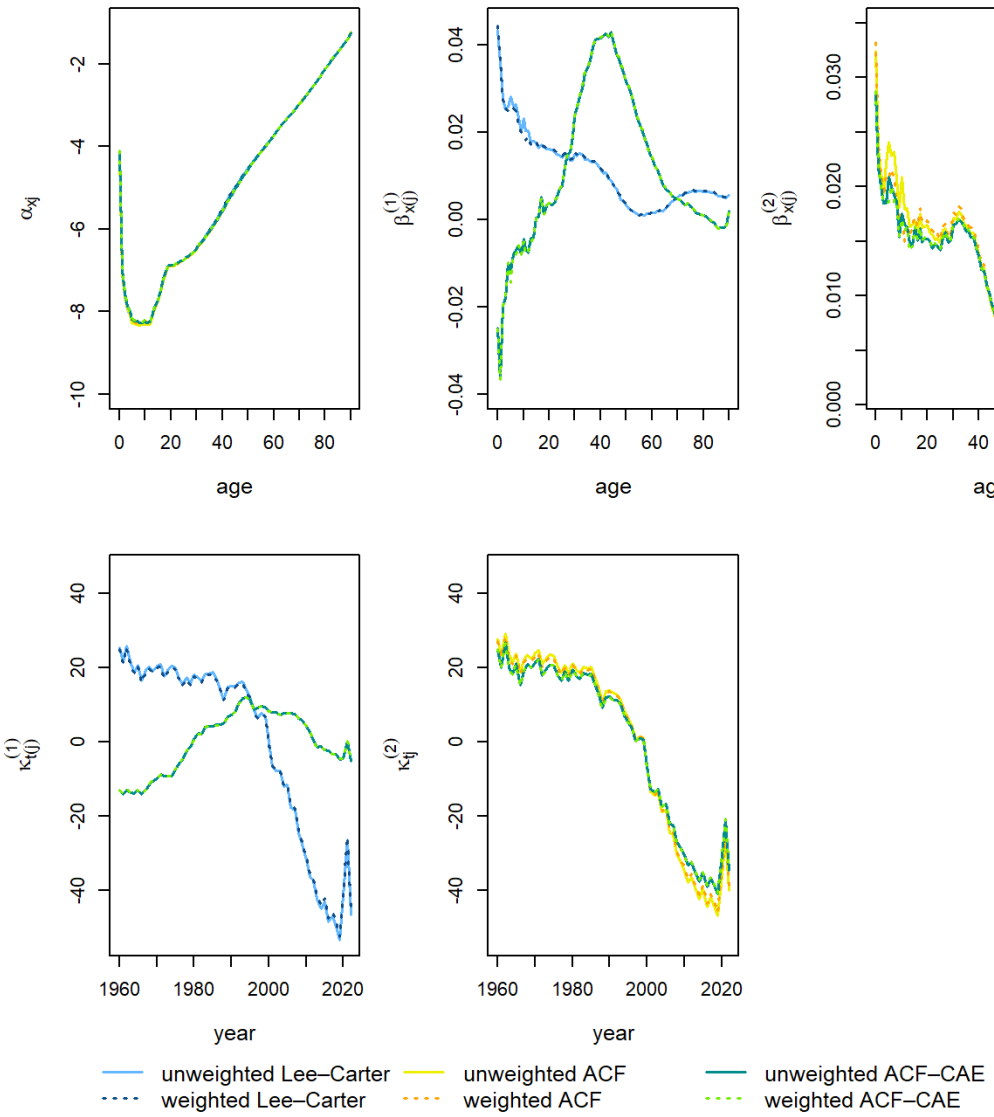
Source: own editing

The fitted parameters

During the analysis, we fitted two types of multi-population model alongside the Lee–Carter model. We also examined weighted versions of all three using Hungarian data. As there were relatively few years and age groups with fewer than five deaths between 1960 and 2022, the parameter values of the weighted and unweighted models differ only slightly. Such a low number of deaths were observed among individuals aged 2 to 18, and the cases were more prevalent among women than men. It is important to add that, considering the ages mentioned, there is a surplus of men. This only reverses after the age of 50. The reason for this is that more boys than girls are born, resulting in an initial surplus of men. However, due to men having a higher mortality rate, this surplus turns into a surplus of women above a certain age. Favourable mortality, defined as fewer than five deaths in a given age group and year, have been observed for women since the mid-1990s and for men since the mid-2000s.

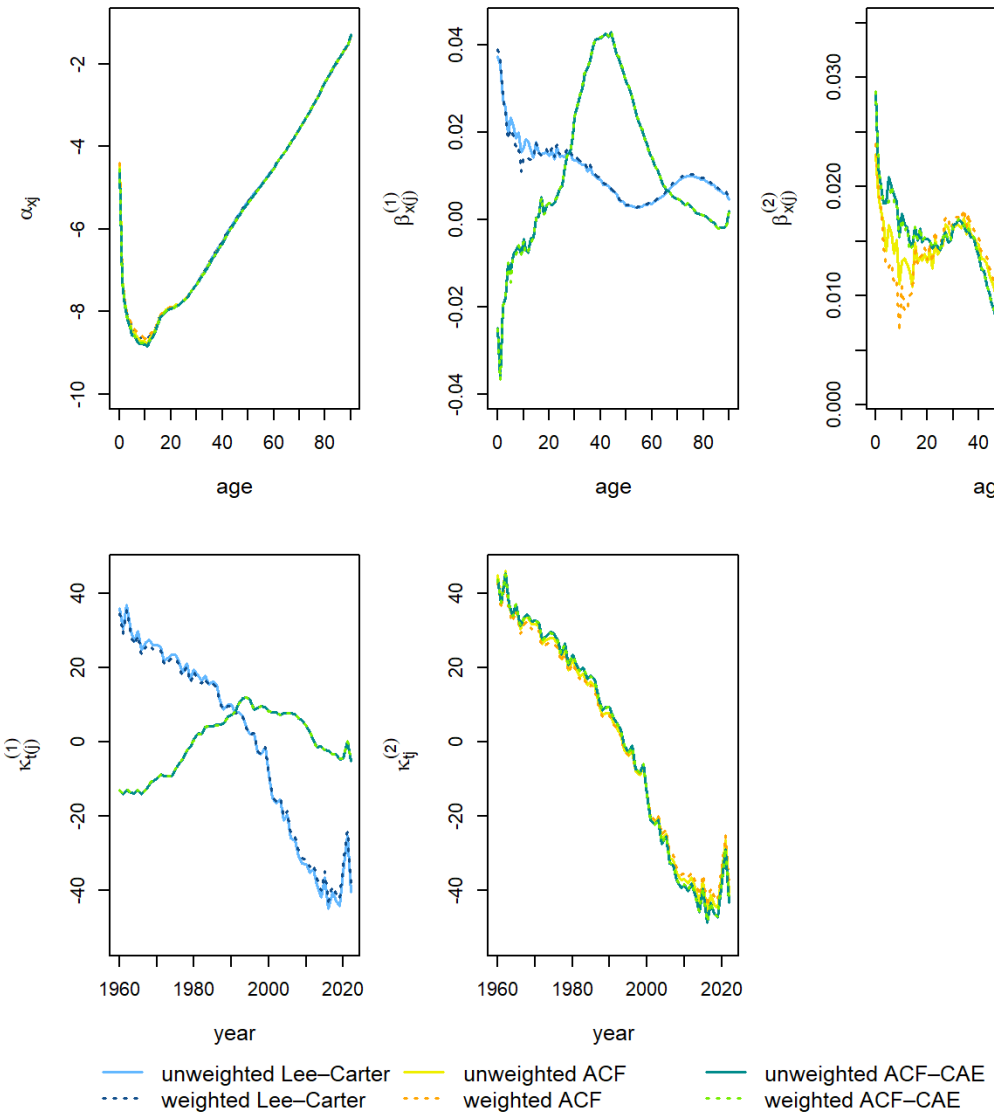
The parameter α_{xj} represents the baseline shape of the log mortality curve and clearly illustrates the difference in mortality between the sexes, with lower values observed for women than for men (see Figures 38–39 and the online Appendix). The β_{xj} parameter values for the Lee–Carter model are positive for both men and women, while the mortality index has been negative in recent decades. This indicates that mortality rates have recently improved for all age groups. In the Lee–Carter model, the parameter $\kappa_{tj}^{(1)}$ shows that the trend of improvement in mortality was more linear for women than for men. Multi-population models provide a clearer picture of mortality trends through their two mortality indices. The variable $\kappa_t^{(1)}$ is a measure of how mortality has changed over time for the whole population. Overall, it can be seen that mortality rates in Hungary deteriorated until the mid-1990s. However, women deviated more from this trend, which offset the unfavourable mortality rate of men to some extent (see the parameters $\kappa_{tj}^{(2)}$ in Figures 38–39). As we explained in Chapter III, women's life expectancy rose steadily during the observed period (excluding the years affected by the COVID-19 pandemic), although this increase was modest until the mid-1990s. Even in the case of two-factor models, mortality indices appear to be predictable based on their trajectory. The second mortality index, which quantifies short- and medium-term deviations from the long-term common trend, is not an irregular component in multi-population models.

Figure 38. The fitted parameters of the weighted and unweighted mortality models for males



Source: own calculation

Figure 39. The fitted parameters of the weighted and unweighted mortality models for females



Source: own calculation

The goodness of fit

The AIC, BIC and MAPE values were calculated for the fitted mortality models as part of the in-sample analysis. According to the AIC and BIC measures, the weighted ACF model was the best fit for both men and women, whereas the unweighted Lee–Carter model was the worst. The former mortality model shows the lowest values, while the latter shows the highest (see Table 22). According to the MAPE values, the unweighted Lee–Carter model also performs less well in terms of goodness of fit. Based on MAPE values, the unweighted ACF model is the best fit for both sexes.

Heatmaps and scatter plots[49] illustrate the magnitude and distribution of the standardized residuals in relation to the dimensions of age, calendar year and year of birth (see the online Appendix). Based on the standardized residuals, it can be concluded that the Lee–Carter model did not fully capture the effects of age, calendar year and cohort. On the heatmaps, the residual values deviate from zero over larger areas in either a positive or negative direction. This is more characteristic of men than women. In the case of the ACF and ACF–CAE models, the diagonal pattern dominates the heatmaps. Based on the scatter plots, these models appear to measure the age- and calendar year-dependent effects more accurately through the added factor. However, the residuals are not independent of the year of birth. All of this confirms the need for combined models.

Table 22. The mortality models' goodness of fit, 1960–2022

Sexes	Measures	unweighted Lee–Carter	unweighted ACF	unweighted ACF–CAE	weighted Lee–Carter	weighted ACF	weighted ACF–CAE
Males	AIC	35,773,942	35,695,354	35,698,399	35,770,565	**35,691,909**	35,694,880
	BIC	35,775,559	35,697,982	35,701,028	35,772,180	**35,694,534**	35,697,505
	MAPE	0.177	**0.103**	0.117	0.181	0.107	0.121
Females	AIC	32,918,361	32,911,775	32,916,640	32,911,176	**32,904,477**	32,909,492
	BIC	32,919,978	32,914,403	32,919,269	32,912,788	**32,907,098**	32,912,112
	MAPE	0.145	**0.129**	0.130	0.156	0.144	0.134

Note: The best values are highlighted in bold.
Source: own calculation

Optimization of the tree-based methods

In the case of the random forests and GBM methods, the hyperparameters need to be optimized. We selected the appropriate hyperparameter setups using expanding window validation. In essence, validation involves using several training periods that are shorter than the base period in order to make predictions for several validation periods. In other words, the training set represents a new fitting period, and the valiadtion set represents the forecast period. Three training periods and three validation periods were defined. The three training periods were as follows: 1960–1990, 1960–2000 and 1960–2010. The corresponding validation periods were: 1991–1999, 2001–2009 and 2011–2019. The

[49] We applied the *StMoMo* package of Villegas et al. (2018) to prepare these figures.

years between 2020 and 2022 were excluded from the validation due to the COVID-19 pandemic. The optimal hyperparameters were selected based on the mean values of the MSE and MAE measures.

As previously mentioned, we applied tree-based machine learning methods in two ways: first, with only the cohort explanatory variable; and second, with both the cohort variable and the COVID-19 dummy variable. In the former case, the random forests method essentially corresponds to bagging, since the variables cannot be randomly selected for each split in each tree. Therefore, $m = p$ (see Chapter I.3). In the second case, the two variables were tried at each split, i.e., the bagging method was also applied. According to these considerations, the number of predictors sampled for spliting was not optimized. It was simply set to one or two.

The following hyperparameters were optimized or set during the expanding window validation process for the random forests model:

- the number of decision trees (*ntree* = 100, 200, 300, 400, 500 or 600),
- the number of predictors sampled for spliting (*mtry* = 1 or 2), and
- the maximum number of terminal nodes (*maxnodes* = 14, 15, 16, 17, 18, 19 or 20).

The tuning parameters that were optimized or set for the GBM are as follows:

- the number of decision trees (*n.trees* = 1,000, 2,000, 3,000, 4,000, 5,000 or 6,000),
- the maximum depth of each tree (*interaction.depth* = 1 or 2),
- the subsampling rate (*bag.fraction* = 0.4, 0.5 or 0.6), and
- the learning rate (*shrinkage* = 0.01, 0.001 or 0.0001).

According to the grid search principle, the MAE and MSE values were examined for all possible combinations of the different tuning parameters (see the online Appendix for more details). The final decision trees were fitted using data from the period between 1960 and 2022. Tables 23–24 show the optimal values for the selected hyperparameters.

Table 23. The optimal hyperparameters for the random forests methods based on the residuals from standard stochastic mortality models

The regression equations	Males			Females		
	ntree	mtry	maxnodes	ntree	mtry	maxnodes
$res_{xtj}^{\text{unweighted Lee–Carter}} \sim cohort$	600	1	14	100	1	18
$res_{xtj}^{\text{weighted Lee–Carter}} \sim cohort$	100	1	18	100	1	18
$res_{xtj}^{\text{unweighted ACF}} \sim cohort$	100	1	15	100	1	14
$res_{xtj}^{\text{weighted ACF}} \sim cohort$	500	1	17	200	1	17
$res_{xtj}^{\text{unweighted ACF–CAE}} \sim cohort$	400	1	15	100	1	15
$res_{xtj}^{\text{weighted ACF–CAE}} \sim cohort$	600	1	19	200	1	14
$res_{xtj}^{\text{unweighted Lee–Carter}} \sim cohort + COVID\ dummy$	600	2	14	100	2	18
$res_{xtj}^{\text{weighted Lee–Carter}} \sim cohort + COVID\ dummy$	100	2	18	100	2	18
$res_{xtj}^{\text{unweighted ACF}} \sim cohort + COVID\ dummy$	100	2	15	100	2	14
$res_{xtj}^{\text{weighted ACF}} \sim cohort + COVID\ dummy$	500	2	17	200	2	17
$res_{xtj}^{\text{unweighted ACF–CAE}} \sim cohort + COVID\ dummy$	400	2	15	100	2	15
$res_{xtj}^{\text{weighted ACF–CAE}} \sim cohort + COVID\ dummy$	600	2	19	200	2	14

Notes: We optimized the *ntree* and *maxnodes* tuning parameters for *mtry* = 1, setting the same values for *mtry* = 2 in the final estimation. We did not consider it important to filter out the impact of the COVID-19 pandemic in this way because this would be taken into account when forecasting mortality indices.
Source: own calculation

Table 24. The optimal hyperparameters for the GBM methods based on the residuals from standard stochastic mortality models

The regression equations	Males				Females			
	n.trees	inter-action. depth	bag. fraction	shrink-age	n.trees	inter-action. depth	bag. fraction	shrink-age
$res_{xtj}^{\text{unweighted Lee–Carter}} \sim cohort$	1,000	1	0.4	0.001	6,000	1	0.6	0.01
$res_{xtj}^{\text{weighted Lee–Carter}} \sim cohort$	5,000	1	0.6	0.0001	1,000	1	0.6	0.0001
$res_{xtj}^{\text{unweighted ACF}} \sim cohort$	6,000	1	0.5	0.0001	2,000	1	0.6	0.001
$res_{xtj}^{\text{weighted ACF}} \sim cohort$	3,000	1	0.4	0.001	4,000	1	0.6	0.001
$res_{xtj}^{\text{unweighted ACF–CAE}} \sim cohort$	3,000	1	0.6	0.0001	1,000	1	0.5	0.0001
$res_{xtj}^{\text{weighted ACF–CAE}} \sim cohort$	4,000	1	0.4	0.0001	1,000	1	0.6	0.0001
$res_{xtj}^{\text{unweighted Lee–Carter}} \sim cohort + COVID\ dummy$	1,000	2	0.4	0.001	6,000	2	0.6	0.01
$res_{xtj}^{\text{weighted Lee–Carter}} \sim cohort + COVID\ dummy$	5,000	2	0.6	0.0001	1,000	2	0.6	0.0001
$res_{xtj}^{\text{unweighted ACF}} \sim cohort + COVID\ dummy$	6,000	2	0.5	0.0001	2,000	2	0.6	0.001
$res_{xtj}^{\text{weighted ACF}} \sim cohort + COVID\ dummy$	3,000	2	0.4	0.001	4,000	2	0.6	0.001
$res_{xtj}^{\text{unweighted ACF–CAE}} \sim cohort + COVID\ dummy$	3,000	2	0.6	0.0001	1,000	2	0.5	0.0001
$res_{xtj}^{\text{weighted ACF–CAE}} \sim cohort + COVID\ dummy$	4,000	2	0.4	0.0001	1,000	2	0.6	0.0001

Note: We optimized the *n.trees*, *bag.fraction*, and *shrinkage* tuning parameters for *interaction.depth* = 1, setting the same values for *interaction.depth* = 2 in the final estimation.
Source: own calculation

Forecasting the mortality indices

Forecasting mortality indices is required for predicting future life expectancy. To this end, a time series analysis was performed. The ADF and KPSS unit root tests indicated that first differencing was necessary for the mortality indices. Several types of ARIMA(p,d,q) models were examined. The parameters were d = 1, and p, q = 0, 1 or 2. The autocorrelation of the residuals of the ARIMA processes was checked using the Ljung–Box Q-statistic and the Breusch–Godfrey test, while normality was examined using the

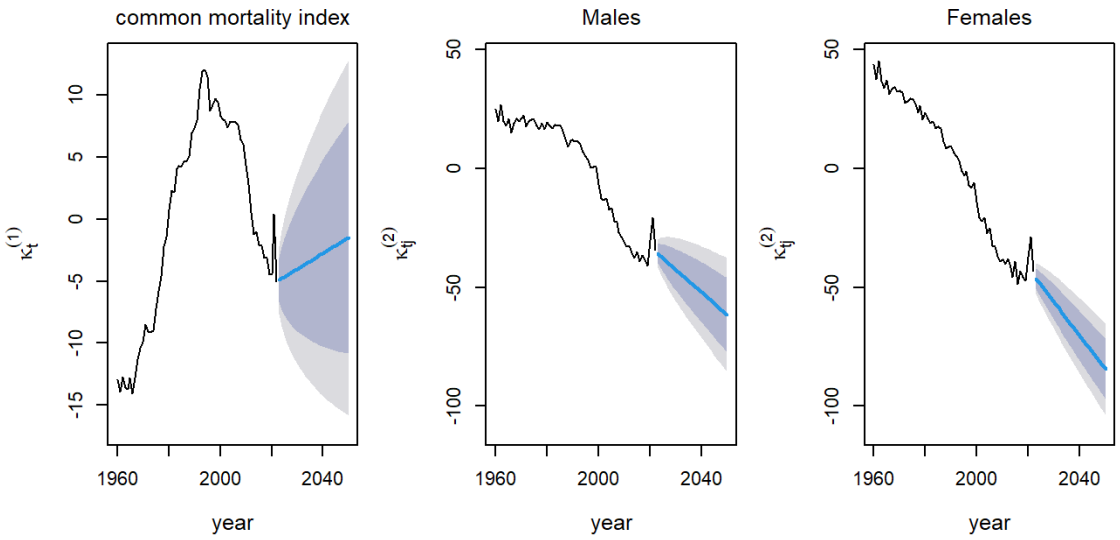
Shapiro–Wilk test. Statistics for the different ARIMA models are available in the online Appendix. The appropriate ARIMA processes were selected based on these results and the AIC and BIC values.[50] Table 25 summarizes the selected ARIMA models according to the mortality model. Figure 40 illustrates the projected mortality indices for the weighted ACF–CAE model. It also shows the 80% and 95% prediction intervals by sex. The projection figures for the other mortality indices can be found in the online Appendix. The impact of the COVID-19 epidemic was filtered out additively across all time series for the years 2020–2021.

Table 25. ARIMA models of the mortality indices

The mortality models	$\kappa_{t(j)}^{(1)}$	$\kappa_{tj}^{(2)}$
unweighted/weighted Lee–Carter	ARIMA(0,1,1) with drift	–
unweighted/weighted ACF	ARIMA(0,1,0) with drift	ARIMA(0,1,1) with drift
unweighted/weighted ACF–CAE	ARIMA(0,1,0) with drift	ARIMA(0,1,1) with drift

Note: ARIMA models do not differ according to sex.
Source: own editing

Figure 40. The common and group-specific mortality indices of the weighted ACF–CAE model



Note: The gray backgrounds represent the 80% and 95% prediction intervals.
Source: own calculation

[50] The formulas for the information criteria are as follows: $AIC = T\ ln(sum\ of\ squared\ residuals) + 2n$ and $BIC = T\ ln(sum\ of\ squared\ residuals) + n\ ln(T)$ where n is the number of estimated parameters (p + q + possible constant term), and T is the number of usable observations (see, e.g., Enders 2014).

The future life expectancy at birth

For standard stochastic mortality models, forecasting the mortality index(es) is sufficient for calculating age-specific mortality rates and life expectancy at birth.[51] In the case of combined models, it is necessary to predict both mortality indices and residuals. This study applied random forests and GBM methods for this purpose. As previously mentioned, 30 different forecasts were created for each sex as part of the study. Life expectancy at birth was predicted up to the year 2050. Thus, machine learning methods were employed to estimate the residuals of mortality models for long-term forecasting purposes. All model forecasts can be found in the online Appendix.

In terms of life expectancy, there is minimal difference, if any, between the weighted and unweighted versions of the same model. This analysis focuses on presenting the results of weighted models. The results of combining the weighted models with the cohort effect estimated by tree-based machine learning methods are shown in Figure 41. According to our calculations, the life expectancy at birth for men was 72.7 years at the beginning of 2022, while for women it was 79.5 years. Before the start of the COVID-19 pandemic, these values were 73.1 and 79.7 years in 2019. According to three combined weighted models, life expectancy for men will reach the 2019 level in 2025. These models are as wollows: „ACF–CAE + cohort effect (random forest)", „ACF–CAE + cohort effect (GBM)", and „ACF–CAE + cohort and COVID effects (GBM)". The „weighted ACF–CAE + cohort effect (random forest)" model predicts the highest value (75.7 years) by 2050, while the weighted Lee–Carter model predicts the lowest value (74 years).

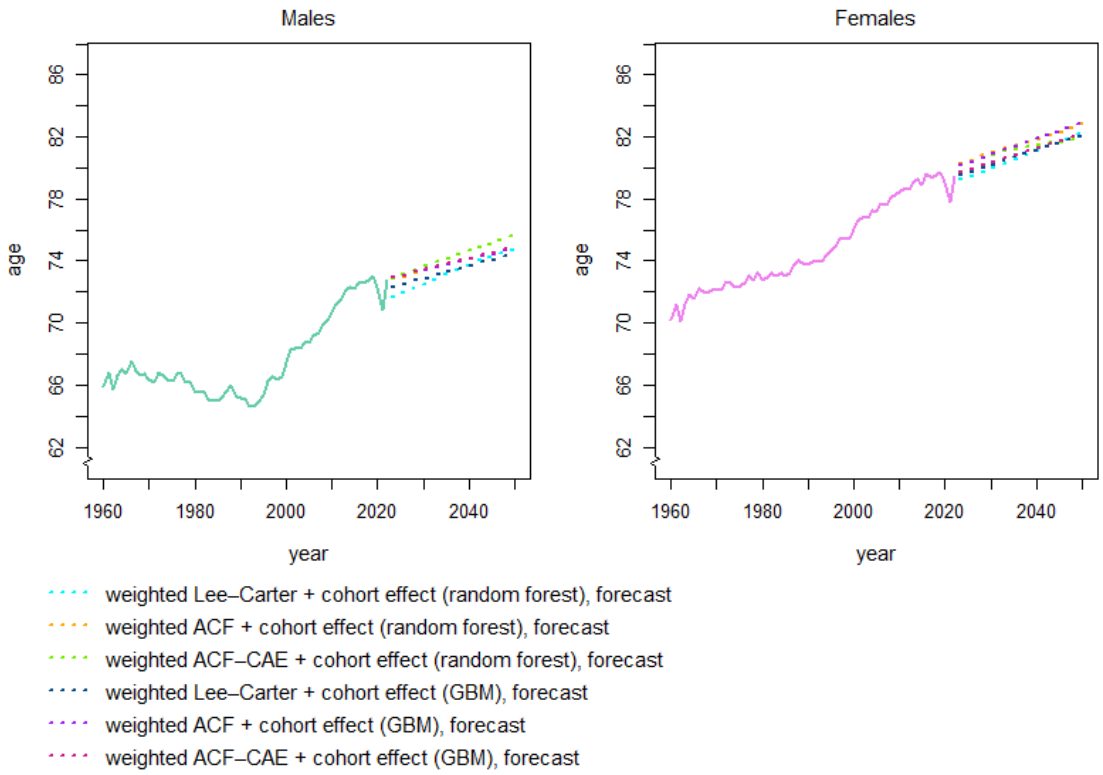
Of the 15 weighted models, only five predict that mortality will not reach the 2019 level in the first year of the forecast (2023) for women. These five models are the Lee–Carter model and its variants combined with machine learning methods. These five models also produced the least favourable projection for men in the first years of the forecast period. By 2023, the „weighted ACF + cohort effect (random forest)" model resulted in the most favourable forecast (80.3 years) for women. The following models only deviated slightly from this: the three other weighted ACF models combined with two machine learning methods, and the two weighted ACF–CAE models combined with random forests. According to the „weighted ACF + cohort effect (GBM)" model, life

[51] In this study, life tables were calculated using the methodology of Chiang (1968) and Eurostat (see https://ec.europa.eu/eurostat/cache/metadata/Annexes/demo_mor_esms_an_1.pdf), except that the group of people aged 90 and over was considered as the oldest cohort instead of people aged 85 and over.

expectancy at birth for women will be the highest in 2050 (83 years), and the lowest according to the weighted Lee–Carter model (81.8 years).

The longevity gap between men and women was 6.8 years in 2022. According to the „weighted ACF–CAE + cohort effect (random forest)" and „weighted ACF–CAE + cohort and COVD effects (random forest)" models, the difference in life expectancy at birth between men and women will be 6.3 years in 2050. All other mortality models predict an increase in the longevity gap to over 7–8 years. Taking all aspects into consideration, the „weighted ACF–CAE + cohort effect (random forest)" model provides the most accurate forecast of life expectancy for both men and women.

Figure 41. Life expectancy at birth for males and females between 1960 and 2050



Source: own calculation

Summary

This study used standard stochastic mortality models and tree-based machine learning methods to analyze long-term time series data. The study examined mortality rates among Hungarian men and women between 1960 and 2022, and made forecasts up to 2050. The three initial models used for mortality forecasting were the ACF and ACF–CAE multi-

population models and the Lee–Carter model. Random forests and GBM methods were applied to forecast residuals from stochastic mortality models in the long term. This approach is potentially novel due to the length of the forecasting horizon, which extends beyond the actual period. The machine learning methods were implemented in two ways. Firstly, we used them to quantify the cohort effect. Secondly, we applied them to model both the cohort effect and the impact of the COVID-19 pandemic, based on the assumption that the mortality index does not fully capture the effects of the epidemic. We also examined the weighted versions of the basic and combined models. A total of 30 forecasts were made regarding the life expectancy of Hungarian men and women.

Based on in-sample analysis, weighted and unweighted ACF models were found to be the best fitted mortality models, while Lee–Carter models performed the worst. Machine learning methods were applied to predict the residuals of standard mortality models. Hyperparameter optimization was performed using expanding window validation to evaluate the results of random forests and GBM methods over three test periods. The forecasts were prepared using the best-trained models. ARIMA models were fitted to predict the mortality indices of stochastic mortality models. The life expectancy results for the weighted and unweighted models were almost the same. Focusing on weighted models, we found that the ACF–CAE model combined with both machine learning methods predicted favourable life expectancy for men. Taking into account the results of the projection and the differences in mortality between the sexes, we conclude that the most appropriate choice is the ACF–CAE model combined with random forests.

The research results clearly show that two-factor multi-population models are more accurate and provide more realistic predictions than the Lee–Carter model. Results also support that models combined with tree-based machine learning methods can provide more accurate predictions of life expectancy, as they also take the cohort effect into account. Using machine learning models is one of the most effective ways of accounting for the cohort effect, as they can capture its nonlinear effects. Based on the results, we do not consider it important to quantify the effect of the pandemic in addition to the cohort effect. Applying outlier filtering to mortality indices can be sufficient.

The study of Bjerre (2022) is the most important research from this field. It is therefore worth comparing our results with this. In contrast with the present study, Bjerre (2022) did not prepare projections beyond the observed period. The researcher found that pure tree-based machine learning models yield the most accurate forecasts compared to standard models, and that random forests and GBM methods do not necessarily improve

the forecast performances of standard stochastic models. Following Bjerre (2022), we fitted pure random forests and pure GBM models to the log-differences of mortality rates directly in order to predict log mortality rates. However, we did not consider pure methods to be suitable for long-term forecasting due to the unrealistic results, so we did not present them. We believe that the combined models performed better than the standard mortality models. However, Bjerre (2022) compared the forecast performance of various models using the model confidence set (MCS) procedure (see Hansen et al. 2011), which was based on different training and validation sets. We did not perform this type of out-of-sample analysis. This could be a further direction for development of the research.

Conclusions

In this thesis, we fitted stochastic mortality models from the Lee–Carter family using Hungarian data to support the development of population forecasting methodology at the HCSO. Our research focused on four main topics, which are as follows: 1) regional mortality projection, 2) analysis and forecast by cause of death, 3) investigation of the time-dependence of age-varying parameters, and incorporation of these parameters into mortality prediction models, also known as rotation, 4) application of tree-based machine learning methods to mortality projection. The four topics are linked by the fact that we used mortality models that can be described within the GAPC model framework. We applied new models to each of the topics that we had not dealt with before. The results of the mortality models were always compared with each other using in-sample and out-of-sample analyses. We used the Lee–Carter model as the main benchmark for all our research topics. The topics are also linked in that we have always fitted multi-population models.

The territorial analysis of mortality is novel in that multi-population models have not yet been applied to Hungarian regional data. We selected the five most well-known multi-population mortality models from the literature. The first research question (Q1) aimed to establish which of the selected models would most accurately forecast long-term male and female mortality at a regional level. Our research revealed that the ACF and ACF–CAE models produced the best results for men. The Lee–Carter model has proven to be appropriate for women. The ACF–CAE model is currently used to forecast mortality at county level in the HCSO as a result of this research. One of the difficulties encountered during the research was that the first years of the forecast horizon coincided with the

period during which the COVID-19 pandemic was subsiding. It was therefore necessary to address the impact of the epidemic. Another difficulty was that there has been no continuous improvement in the mortality trend among Hungarian men over the past few decades. Therefore, it is difficult to forecast mortality improvement in their case.

One of the innovations in analyzing the main causes of mortality in Hungary was the development of multi-population models, which were created by clustering the mortality indices of single-population models. We used the dynamic time warping algorithm to calculate the distance between the cause-specific mortality indices. We fitted seven single-population mortality models and then developed multi-population versions of each one. Another new approach to this topic involved using multi-population models to examine mortality by the main causes of death, which are not independent of each other. When selecting single-population models, we opted for ones that had not previously been applied to Hungarian data in our research, and that would enable us to become more familiar with models featuring predefined age-dependent parameters. In addition, we used models that assumed a Poisson distribution of deaths and also fitted models of mortality with a Binomial distribution. Another outcome of the research is that we have developed multi-population models that have not yet been discussed in international literature. Of the fitted models, the reduced Plat model with three factors and its multi-population version can be highlighted as a well-performing model. With regard to the research question (Q2) relating to this topic, we found that multi-population models may be an appropriate way of capturing the relationship between causes of death. This is because they examine similar causes of death in terms of mortality trends and forecast common trends.

Since the rate of improvement in mortality may vary among different age groups, we considered it important to examine rotation. We analyzed the time-dependence of the age-varying parameters using least squares regression. The innovative aspect of this research is that we have incorporated rotation into the forecast, even for multi-population models where age-varying parameters have not yet been rotated in the international literature. In our analysis by sex, we opted for augmented mortality models, as the ACF model produced accurate regional forecasts. We extended this model by adding an extra factor, which we assumed would improve the goodness of fit. Of the rotated and nonrotated models, the ACF model with three factors was found to be the most suitable for forecasting. However, given that life expectancy at birth is increasing at an ever-faster rate, the rotated three-factor ACF model cannot be used for long-term forecasting. The

research used a rotation method that is better suited to short- or medium-term forecasts than long-term ones. Regarding research question (Q3), we found that considering the rotation of age-dependent parameters could result in more accurate predictions. In particular, it is worth incorporating the rotation of the parameter that describes the baseline shape of mortality into the forecast. This could serve as an alternative to the Lee–Miller (2001) method during the initial years of the forecast, thereby avoiding jump-off bias.

When applying standard stochastic mortality models, we assumed a linear relationship between log mortality rates and time- or age-dependent parameters. Machine learning methods can be used to capture nonlinear relationships. We primarily exploited this advantage when applying tree-based machine learning methods to model cohort effects. We applied machine learning methods to the residuals of standard mortality models, analyzing the data by sex. During the research, we selected two augmented multi-population mortality models. These had an appropriate fit regarding other research topics. The main aim was to incorporate cohort effects into the selected models. In response to the fourth research question (Q4), we hypothesized that the combined models would produce more accurate predictions. This hypothesis was confirmed by the research results. In comparison to other topics, the fact that we prepared weighted versions of mortality models was novel. Another novelty in this research was that the length of the forecasting horizon extended beyond the observed period. In addition to the aforementioned results and innovations, we developed our own numerical programming to fit all mortality models. This enables us to adjust the models' parameters according to our own criteria.

References

Aburto–van Raalte (2018): Aburto, José Manuel – van Raalte, Alyson, 2018, Lifespan dispersion in times of life expectancy fluctuation: The case of Central and Eastern Europe. *Demography* **55(6)**, pp 2071–2096. https://doi.org/10.1007/s13524-018-0729-9

Arnold–Glushko (2021): Arnold (-Gaille), Séverine – Glushko, Viktoriya, 2021, Cause-specific mortality rates: Common trends and differences. *Insurance: Mathematics and Economics* **99**, pp 294–308. https://doi.org/10.1016/j.insmatheco.2021.03.027

Arnold–Sherris (2013): Arnold (-Gaille), Séverine – Sherris, Michael, 2013, Forecasting mortality trends allowing for cause-of-death mortality dependence. *North American Actuarial Journal* **17(4)**, pp 273–282. https://doi.org/10.1080/10920277.2013.838141

Arnold–Sherris (2015): Arnold (-Gaille), Séverine – Sherris, Michael, 2015, Causes-of-death mortality: What do we know on their dependence? *North American Actuarial Journal* **19(2)**, pp 116–128. https://doi.org/10.1080/10920277.2015.1011279

Bajkó et al. (2015): Bajkó, Attila – Maknics, Anita – Tóth, Krisztián – Vékás, Péter, 2015, A magyar nyugdíjrendszer fenntarthatóságáról. *Közgazdasági Szemle* **62(12)**, pp 1229–1257. https://doi.org/10.18414/KSZ.2015.12.1229

Bálint (2011): Bálint, Lajos, 2011, A születéskor várható élettartam nemek szerinti térbeli különbségei. *Területi Statisztika* **51(4)**, pp 386–404. https://www.ksh.hu/statszemle_archive/terstat/2011/2011_04/ts2011_04_04.pdf

Bálint (2016): Bálint, Lajos, 2016, Mennyire illeszkedik a magyar halandóság alakulása az epidemiológiai átmenet elméletéhez? *Demográfia* **59(1)**, pp 5–57. https://doi.org/10.21543/Dem.59.1.1

Baran et al. (2007): Baran, Sándor – Gáll, József – Ispány, Márton – Pap, Gyula, 2007, Forecasting Hungarian mortality rates using the Lee–Carter method. *Acta Oeconomica* **57(1)**, pp 25–38. https://doi.org/10.1556/AOecon.57.2007.1.3

Bellman–Kalaba (1959): Bellman, Richard – Kalaba, Robert E., 1959, On adaptive control processes. *IRE Transactions on Automatic Control* **4(2)**, pp 1–9. https://doi.org/10.1109/TAC.1959.1104847

Bjerre (2022): Bjerre, Dorethe Skovgaard, 2022, Tree-based machine learning methods for modeling and forecasting mortality. *ASTIN Bulletin* **52(3)**, pp 765–787. https://doi.org/10.1017/asb.2022.11

Booth et al. (2002): Booth, Heather – Maindonald, John – Smith, Len, 2002, Applying Lee-Carter under conditions of variable mortality decline. *Population Studies* **56(3)**, pp 325–336. https://doi.org/10.1080/00324720215935

Breiman (1996a): Breiman, Leo, 1996, Bagging predictors. *Machine Learning* **24**, pp 123–140. https://doi.org/10.1007/BF00058655

Breiman (1996b): Breiman, Leo, 1996, Heuristics of instability and stabilization in model selection. *The Annals of Statistics* **24(6)**, pp 2350–2383.
http://www.jstor.org/stable/2242688

Breiman (2001): Breiman, Leo, 2001, Random forests. *Machine Learning* **45**, pp 5–32.
https://doi.org/10.1023/A:1010933404324

Breiman et al. (1984): Breiman, Leo – Friedman, Jerome H. – Stone, Charles J. – Olshen, Richard A., 1984, Classification and regression trees. *Wadsworth International Group*, Belmont.
https://doi.org/10.1201/9781315139470

Brillinger (1986): Brillinger, David R., 1986, The natural variability of vital rates and associated statistics. *Biometrics* **42(4)**, pp 693–734.
https://doi.org/10.2307/2530689

Brouhns et al. (2002): Brouhns, Natacha – Denuit, Michel – Vermunt, Jeroen K., 2002, A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* **31(3)**, pp 373–393.
https://doi.org/10.1016/S0167-6687(02)00185-3

Butt et al. (2014): Butt, Zoltan – Haberman, Steven – Shang, Han Lin, 2014, The ilc package.
https://rdrr.io/cran/ilc/

Cairns et al. (2006): Cairns, Andrew J. G. – Blake, David – Dowd, Kevin, 2006, A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance* **73(4)**, pp 687–718.
https://doi.org/10.1111/j.1539-6975.2006.00195.x

Cairns et al. (2011): Cairns, Andrew J. G. – Blake, David – Dowd, Kevin – Coughlan, Guy D. – Epstein, David – Khalaf-Allah, Marwa, 2011, Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics* **48(3)**, pp 355–367.
https://doi.org/10.1016/j.insmatheco.2010.12.005

Cairns et al. (2009): Cairns, Andrew J. G. – Blake, David – Dowd, Kevin – Coughlan, Guy D. – Epstein, David – Ong, Alen – Balevich, Igor, 2009, A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal* **13(1)**, pp 1–35.
https://doi.org/10.1080/10920277.2009.10597538

Cairns et al. (2020): Cairns, Andrew J. G. – Blake, David – Kessler, Amy – Kessler, Marsha, 2020, The impact of COVID-19 on future higher-age mortality. *Pensions Institute Working Paper* PI-2007
http://www.pensions-institute.org/wp-content/uploads/wp2007.pdf

Carriere (1994): Carriere, Jacques F., 1994, Dependent decrement theory. *Transactions of Society of Actuaries* **46**, pp 45–65.

Carter–Lee (1992): Carter, Lawrence R. – Lee, Ronald D., 1992, Modeling and forecasting US sex differentials in mortality. *International Journal of Forecasting* **8(3)**, pp 393–411.

Chiang (1968): Chiang, Chin Long, 1968, Introduction to stochastic processes in biostatistics. *Wiley*, New York.

Coelho–Nunes (2011): Coelho, Edviges – Nunes, Luis C., 2011, Forecasting mortality in the event of a structural change. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **174(3)**, pp 713–736.
https://doi.org/10.1111/j.1467-985X.2010.00687.x

Csiszár (2022): Csiszár, Dorottya, 2022, Elkerülhető halálozás vizsgálata magyar és ír adatokon. PCLM, P-spline és Lee-Carter modell alkalmazása. *Biztosítás és Kockázat* **9(1–2)**, pp 12–43.
https://doi.org/10.18530/BK.2022.1-2.12

Currie (2006): Currie, Iain D., 2006, Smoothing and forecasting mortality rates with P-splines. Technical Report, Heriot-Watt University.
https://www.actuaries.org.uk/system/files/documents/pdf/currie.pdf

Danesi et al. (2015): Danesi, Ivan Luciano – Haberman, Steven – Millossovich, Pietro, 2015, Forecasting mortality in subpopulations using Lee–Carter type models: A comparison. *Insurance: Mathematics and Economics* **62**, pp 151–161.
http://dx.doi.org/10.1016/j.insmatheco.2015.03.010

Deprez et al. (2017): Deprez, Philippe – Shevchenko, Pavel V. – Wüthrich, Mario V., 2017, Machine learning techniques for mortality modeling. *European Actuarial Journal* **7**, pp 337–352.
https://doi.org/10.1007/s13385-017-0152-4

Dimitrova et al. (2013): Dimitrova, Dimitrina S. – Haberman, Steven – Kaishev, Vladimir K., 2013, Dependent competing risks: cause elimination and its impact on survival. *Insurance: Mathematics and Economics* **53(2)**, pp 464–477.
https://doi.org/10.1016/j.insmatheco.2013.07.008

Dowd et al. (2020): Dowd, Kevin – Cairns, Andrew J. G. – Blake, David, 2020, CBDX: A workhorse mortality model from the Cairns–Blake–Dowd family. *Annals of Actuarial Science* **14(2)**, pp 445–460.
https://doi.org/10.1017/S1748499520000159

Enchev et al. (2017): Enchev, Vasil – Kleinow, Torsten – Cairns, Andrew J. G., 2017, Multi-population mortality models: Fitting, forecasting and comparisons. *Scandinavian Actuarial Journal* **2017(4)**, pp 319–342.
https://doi.org/10.1080/03461238.2015.1133450

Enders (2014): Enders, Walter, 2014, Applied econometric time series. *Wiley*, 4th edition, New York.

Friedman (2001): Friedman, Jerome H., 2001, Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* **29(5)**, pp 1189–1232.
https://doi.org/10.1214/aos/1013203451

Giorgino (2009): Giorgino, Toni, 2009, Computing and visualizing dynamic time warping alignments in R: The dtw package. *Journal of Statistical Software* **31(7)**, pp 1–24.
https://doi.org/10.18637/jss.v031.i07

Gogola–Vékás (2020): Gogola, Ján – Vékás, Péter, 2020, Élettartam-kockázat Csehországban és Magyarországon. *Biztosítás és Kockázat* **7(3–4)**, pp 14–26.
https://doi.org/10.18530/BK.2020.3-4.14

Goldstein–Lee (2024): Goldstein, Joshua R. – Lee, Ronald D., 2024, Life expectancy reversals in low-mortality populations. *Population and Development Review* **50(2)**, pp 437–459.
https://doi.org/10.1111/padr.12619

Hansen et al. (2011): Hansen, Peter R. – Lunde, Asger – Nason, James M., 2011, The model confidence set. *Econometrica* **79(2)**, pp 453–497.
https://doi.org/10.3982/ECTA5771

Harris et al. (2009): Harris, David – Harvey, David I. – Leybourne, Stephen J. – Taylor, A. M. Robert, 2009, Testing for a unit root in the presence of a possible break in trend. *Econometric Theory* **25(6)**, pp 1545–1588.
https://doi.org/10.1017/S0266466609990259

Harvey et al. (2009): Harvey, David I. – Leybourne, Stephen J. – Taylor, A. M. Robert, 2009, Simple, robust, and powerful tests of the breaking trend hypothesis. *Econometric Theory* **25(4)**, pp 995–1029.
https://doi.org/10.1017/S0266466608090385

Hastie et al. (2009): Hastie, Trevor – Tibshirani, Robert – Friedman, Jerome, 2009, The elements of statistical learning. *Springer-Verlag*, 2nd edition, New York.

HCSO (1955–1964): Hungarian Central Statistical Office, 1955–1964, Magyarország népesedése. *Központi Statisztikai Hivatal*, Budapest.

HCSO (1965–present): Hungarian Central Statistical Office, 1965–present, Demográfiai évkönyv. *Központi Statisztikai Hivatal*, Budapest.

HCSO (1990): Hungarian Central Statistical Office, 1990, Korrigált népességszámok megyénként, 1980–1990. *Központi Statisztikai Hivatal*, Budapest.

HCSO (2001): Hungarian Central Statistical Office, 2001, A 2001. február 1-jei népszámlálás végleges adatai alapján korrigált 1990–2001. évi továbbszámított népességszámok. Munkaanyag. *Központi Statisztikai Hivatal*, Budapest.

HCSO (2023): Hungarian Central Statistical Office, 2023, Demographic Yearbook, 2022 *Hungarian Central Statistical Office*, Budapest.

Hrzic et al. (2020): Hrzic, Rok – Vogt, Tobias – Janssen, Fanny – Brand, Helmut, 2020, Mortality convergence in the enlarged European Union: A systematic literature review. *European Journal of Public Health* **30(6)**, pp 1108–1115.
https://doi.org/10.1093/eurpub/ckaa038

Hunt–Blake (2014): Hunt, Andrew – Blake, David, 2014, A general procedure for constructing mortality models. *North American Actuarial Journal* **18(1)**, pp 116–138.
https://doi.org/10.1080/10920277.2013.852963

Hunt–Blake (2015): Hunt, Andrew – Blake, David, 2015, On the structure and classification of mortality models. *Pensions Institute Working Paper* PI-1506
http://www.pensions-institute.org/wp-content/uploads/2019/workingpapers/wp1506.pdf

Hunt–Blake (2020): Hunt, Andrew – Blake, David, 2020, Identifiability in age/period mortality models. *Annals of Actuarial Science* **14(2)**, pp 461–499.
https://doi.org/10.1017/S1748499520000111

Hyndman (2013): Hyndman, Rob J. – Booth, Heather – Yasmeen, Farah, 2013, Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography* **50(1)**, pp 261–283.
https://doi.org/10.1007/s13524-012-0145-5

James et al. (2013): James, Gareth – Witten, Daniela – Hastie, Trevor – Tibshirani, Robert, 2013, An introduction to statistical learning with applications in R. *Springer Science and Business Media*, New York. ISBN 978-1-4614-7137-0
https://doi.org/10.1007/978-1-4614-7138-7

Jarner–Jallbjørn (2020): Jarner, Søren F. – Jallbjørn, Snorre, 2020, Pitfalls and merits of cointegration-based mortality models. *Insurance: Mathematics and Economics* **90**, pp 80–93.
https://doi.org/10.1016/j.insmatheco.2019.10.005

Jasilionis et al. (2023): Jasilionis, Domantas – Meslé, France – Vallin, Jacques, 2023, Is East-West life expectancy gap narrowing in the enlarged European Union? *Comparative Population Studies* **48**, pp 523–552.
https://doi.org/10.12765/CPoS-2023-20

Józan (2006): Józan, Péter, 2006, A new Hungarian revolution. *Circulation* **113(5)**, F20
https://www.ahajournals.org/doi/pdf/10.1161/circ.113.5.f17

Józan (2008): Józan, Péter, 2008, Válság és megújulás a második világháború utáni epidemiológiai fejlődésben Magyarországon. *MTA Társadalomkutató Központ*, Budapest.

Józan (2009): Józan, Péter, 2009, Csökkenő kardiovaszkuláris mortalitás, javuló életkilátások, új epidemiológiai korszak kezdete Magyarországon. *IME Interdiszciplináris Magyar Egészségügy* **8(4)**, pp 21–25.
https://www.imeonline.hu/article.php?article=2009._VIII./4/csokkeno_kardiovaszkularis_mortalitas_javulo_eletkilatasok_uj_epidemiologiai_korszak_kezdete_magyarorszagon

Józan (2012): Józan, Péter, 2012, Rendszerváltozás és epidemiológiai korszakváltás Magyarországon. *Orvosi Hetilap* **153(17)**, pp 662–677.
https://doi.org/10.1556/oh.2012.29344

Kaishev et al. (2007): Kaishev, Vladimir K. – Dimitrova, Dimitrina S. – Haberman, Steven, 2007, Modelling the joint distribution of competing risks survival times using copula functions. *Insurance: Mathematics and Economics* **41(3)**, pp 339–361.
https://doi.org/10.1016/j.insmatheco.2006.11.006

Kleinow (2015): Kleinow, Torsten, 2015, A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics* **63(C)**, pp 147–152.
https://doi.org/10.1016/j.insmatheco.2015.03.023

Lee (2000): Lee, Ronald D., 2000, The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal* **4(1)**, pp 80–93.
https://doi.org/10.1080/10920277.2000.10595882

Lee–Carter (1992): Lee, Ronald D. – Carter, Lawrence R., 1992, Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association* **87(419)**, pp 659–671.
https://doi.org/10.2307/2290201

Lee–Miller (2001): Lee, Ronald D. – Miller, Timothy, 2001, Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography* **38(4)**, pp 537–549.
https://doi.org/10.1353/dem.2001.0036

Lee–Nault (1993): Lee, Ronald D. – Nault, Francois, 1993, Modeling and forecasting provincial mortality in Canada. Presented at the World Congress of the International Union for Scientific Study of Population. Montréal, Canada.

Levantesi–Nigri (2020): Levantesi, Susanna – Nigri, Andrea, 2020, A random forest algorithm to improve the Lee–Carter mortality forecasting: impact on q-forward. *Soft Computing* **24**, pp 8553–8567. https://doi.org/10.1007/s00500-019-04427-z

Levantesi–Pizzorusso (2019): Levantesi, Susanna – Pizzorusso, Virginia, 2019, Application of machine learning to mortality modeling and forecasting. *Risks* **7(1)**, pp 1–19. https://doi.org/10.3390/risks7010026

Li et al. (2019): Li, Han – Li, Hong – Lu, Yang – Panagiotelis, Anastasios, 2019, A forecast reconciliation approach to cause-of-death mortality modeling. *Insurance: Mathematics and Economics* **86(C)**, pp 122–133. https://doi.org/10.1016/j.insmatheco.2019.02.011

Li–Lu (2018): Li, Hong – Lu, Yang, 2018, Modeling cause-of-death mortality using hierarchical Archimedean copula. *Scandinavian Actuarial Journal* **2019(3)**, pp 247–272. https://doi.org/10.1080/03461238.2018.1546224

Li (2013): Li, Jackie, 2013, A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies* **67(1)**, pp 111–126. https://doi.org/10.1080/00324728.2012.689316

Li et al. (2016): Li, Jackie – Tickle, Leonie – Parr, Nick, 2016, A multi-population evaluation of the Poisson common factor model for projecting mortality jointly for both sexes. *Journal of Population Research* **33(4)**, pp 333–360. https://doi.org/10.1007/s12546-016-9173-0

Li et al. (2011): Li, Johnny Siu-Hang – Chan, Wai-Sum – Cheung, Siu-Hung, 2011, Structural changes in the Lee-Carter mortality indexes: Detection and implications. *North American Actuarial Journal* **15(1)**, pp 13–31. https://doi.org/10.1080/10920277.2011.10597607

Li–Gerland (2011): Li, Nan – Gerland, Patrick, 2011, Modifying the Lee-Carter method to project mortality changes up to 2100. Presented at the 76th Annual Meeting of the Population Association of America. Washington, D.C., USA. https://population.un.org/wpp/publications/Files/Li_2011_Modifying%20the%20Lee-Carter%20method%20to%20project%20mortality%20changes%20up%20to%202100.pdf

Li–Lee (2005): Li, Nan – Lee, Ronald D., 2005, Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography* **42(3)**, pp 575–594. https://www.jstor.org/stable/4147363

Li et al. (2013): Li, Nan – Lee, Ronald D. – Gerland, Patrick, 2013, Extending the Lee-Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography* **50(6)**, pp 2037–2051. https://doi.org/10.1007/s13524-013-0232-2

Lourés–Cairns (2021): Lourés, Cristian Redondo – Cairns, Andrew J. G., 2021, Cause of death specific cohort effects in U.S. mortality. *Insurance: Mathematics and Economics* **99**, pp 190–199. https://doi.org/10.1016/j.insmatheco.2021.03.026

Májer–Kovács (2011): Májer, István – Kovács, Erzsébet, 2011, Élettartam-kockázat – a nyugdíjrendszerre nehezedő egyik teher. *Statisztikai Szemle* **89(7–8)**, pp 790–812.
https://www.ksh.hu/statszemle_archive/all/2011/2011_07-08/2011_07-08_790.pdf

Marino et al. (2021): Marino, Mario – Levantesi, Susanna – Nigri, Andrea, 2021, Deepening Lee–Carter for longevity projections with uncertainty estimation. *arXiv*, arXiv: 2103.10535.
https://doi.org/10.48550/arXiv.2103.10535

Meslé–Vallin (2002): Meslé, France – Vallin, Jacques, 2002, Mortality in Europe: The divergence between East and West. *Population* **57(1)**, pp 157–197.
https://doi.org/10.2307/3246630

Nigri et al. (2019): Nigri, Andrea – Levantesi, Susanna – Marino, Mario – Scognamiglio, Salvatore – Perla, Francesca, 2019, A deep learning integrated Lee–Carter model. *Risks* **7(1)**, pp 1–16.
https://doi.org/10.3390/risks7010033

Obádovics–Tóth (2021): Obádovics, Csilla – Tóth, Csaba G., 2021, A népesség szerkezete és jövője. In Monostori, Judit – Őri, Péter – Spéder, Zsolt (szerk.), *Demográfiai Portré 2021. Jelentés a magyar népesség helyzetéről*. *KSH Népességtudományi Kutatóintézet*, Budapest, pp 251–275.
https://www.demografia.hu/kiadvanyokonline/index.php/demografiaiportre/article/view/2837/2727

Obádovics–Tóth (2023): Obádovics, Csilla – Tóth, Csaba G., 2023, A magyarországi régiók népességének előreszámítása 2050-ig. *Statisztikai Szemle* **101(9)**, pp 763–792.
https://doi.org/10.20311/stat2023.09.hu0763

Perla et al. (2021): Perla, Francesca – Richman, Ronald – Scognamiglio, Salvatore, 2021, Time-series forecasting of mortality rates using deep learning. *Scandinavian Actuarial Journal* **2021(7)**, pp 572–598.
https://doi.org/10.1080/03461238.2020.1867232

Petneházi–Gáll (2019): Petneházi, Gábor – Gáll, József, 2019, Mortality rate forecasting: can recurrent neural networks beat the Lee-Carter model? *arXiv*, arXiv: 1909.05501.
https://doi.org/10.48550/arXiv.1909.05501

Petneházi–Gáll (2023): Petneházi, Gábor – Gáll, József, 2023, Testing the Lee-Carter model on Hungarian mortality data. *Acta Oeconomica* **73(1)**, pp 171–182.
https://doi.org/10.1556/032.2023.00010

Piveteau (2020): Piveteau, Samuel, 2020, Mortality clustering with the K-Lee-Carter. Presented at the 3rd Human Mortality Database Users' Virtual Conference.
https://www.scor.com/sites/default/files/2020-12/Piveteau_HMD_MortalityClustering.pdf

Piveteau (2021): Piveteau, Samuel, 2021, Modélisation de la mortalité par cause de décès. Gestion et management. *Université de Lyon*, Français.
https://theses.hal.science/tel-03499914

Plat (2009): Plat, Richard, 2009, On stochastic mortality modeling. *Insurance: Mathematics and Economics* **45(3)**, pp 393–404.
https://doi.org/10.1016/j.insmatheco.2009.08.006

Renshaw–Haberman (2003): Renshaw, Arthur – Haberman, Steven, 2003, Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics* **33(2)**, pp 255–272.
https://doi.org/10.1016/S0167-6687(03)00138-0

Renshaw–Haberman (2006): Renshaw, Arthur – Haberman, Steven, 2006, A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* **38(3)**, pp 556–570.
https://doi.org/10.1016/j.insmatheco.2005.12.001

Richman (2018): Richman, Ronald, 2018, AI in Actuarial Science. *SSRN Electronic Journal*.
http://dx.doi.org/10.2139/ssrn.3218082

Richman–Wüthrich (2019): Richman, Ronald – Wüthrich, Mario V., 2018, Lee and Carter go machine learning: Recurrent neural networks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3441030

Richman–Wüthrich (2021): Richman, Ronald – Wüthrich, Mario V., 2021, A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science* **15(2)**, pp 346–366.
https://doi.org/10.1017/S1748499519000071

Sardá-Espinosa (2017): Sardá-Espinosa, Alexis, 2017, Comparing time-series clustering algorithms in R using the dtwclust package. Manuscript, pp 1–45.
https://cran.r-project.org/web/packages/dtwclust/vignettes/dtwclust.pdf

Sardá-Espinosa (2019): Sardá-Espinosa, Alexis, 2019, Time-series clustering in R using the dtwclust package. *The R Journal* **11(1)**, pp 22–43.
https://doi.org/10.32614/RJ-2019-023

Schnürch–Korn (2022): Schnürch, Simon – Korn, Ralf, 2022, Point and interval forecasts of death rates using neural networks. *ASTIN Bulletin* **52(1)**, pp 333–360.
https://doi.org/10.1017/asb.2021.34

Scognamiglio (2022): Scognamiglio, Salvatore, 2022, Longevity risk analysis: applications to the Italian regional data. *Quantitative Finance and Economics* **6(1)**, pp 138–157.
https://doi.org/10.3934/QFE.2022006

Simpach–Arltova (2016): Simpach, Ondrej – Arltova, Marketa: An increasing of prediction power of the Lee-Carter model: The case of Czech and Spanish age-specific fertility rates' forecasting. *University of Economics Prague*, Faculty of Informatics and Statistics, Department of Statistics and Probability.

Sobreira–Nunes (2016): Sobreira, Nuno – Nunes, Luis C., 2016, Tests for multiple breaks in the trend with stationary or integrated shocks. *Oxford Bulletin of Economics and Statistics* **78(3)**, pp 394–411.
https://doi.org/10.1111/obes.12116

Szentkereszti–Vékás (2022): Szentkereszti, Gábor – Vékás, Péter, 2022, Magyar halandósági ráták előrejelzése visszacsatolt neurális hálózatokkal. *Statisztikai Szemle* **100(10)**, pp 905–922.
https://doi.org/10.20311/stat2022.10.hu0905

Szentkereszti–Vékás (2024): Szentkereszti, Gábor – Vékás, Péter, 2024, A visegrádi országok mortalitási rátáinak előrejelzése neurális hálózatokkal. *Szigma* **55(2–3)**, pp 293–312.
https://doi.org/10.15170/SZIGMA.55.1242

Thatcher et al. (1998): Thatcher, A. Roger – Kannisto, Väinö – Vaupel, James W., 1998, The force of mortality at ages 80 to 120. *Odense Monographs on Population Aging Series*, No. 5, Odense University Press.
https://www.demogr.mpg.de/Papers/Books/Monograph5/ForMort.htm

Tóth (2021a): Tóth, Csaba G., 2021, Többlethalandóság a koronavírus-járvány miatt Magyarországon 2020-ban. *Korfa* **2021(2)**, pp 1–4.
https://demografia.hu/kiadvanyokonline/index.php/korfa/article/view/2812/2700

Tóth (2021b): Tóth, Csaba G., 2021, Multi-population models to handle mortality crises in forecasting mortality: A case study from Hungary. *Society and Economy* **43(2)**, pp 128–146.
https://doi.org/10.1556/204.2021.00007

Tóth (2022a): Tóth, Csaba G., 2022, Másfél év pandémia Magyarországon: Mérséklődő különbségek a regionális és korspecifikus többlethalandóságban. KRTK-KTI Műhelytanulmányok 2022/04, *Közgazdaság- és Regionális Tudományi Kutatóközpont*, Közgazdaság-tudományi Intézet, Budapest.
https://kti.krtk.hu/wp-content/uploads/2022/01/CERSIEWP202204.pdf

Tóth (2022b): Tóth, Csaba G., 2022, Narrowing the gap in regional and age-specific excess mortality during the COVID-19 in Hungary. *Eastern Journal of European Studies* **13(1)**, pp 185–207.
https://doi.org/10.47743/ejes-2022-0109

Vallin–Meslé (2004): Vallin, Jacques – Meslé, France, 2004, Convergences and divergences in mortality. A new approach to health transition. *Demographic Research* **S2**, pp 11–44.
https://doi.org/10.4054/DemRes.2004.S2.2

Vallin–Meslé (2009): Vallin, Jacques – Meslé, France, 2009, The segmented trend line of highest life expectancies. *Population and Development Review* **35(1)**, pp 159–187.
https://doi.org/10.1111/j.1728-4457.2009.00264.x

Varga (2023): Varga, Livia, 2023, Fitting and forecasting multi-population mortality models based on Hungarian regional data. *Regional Statistics* **13(5)**, pp 863–898.
https://doi.org/10.15196/RS130504

Varga (2024): Varga, Livia, 2024, Rotation of the age-varying parameters in multi-population mortality models. *Scientia et Securitas* **4(4)**, pp 375–391.
https://doi.org/10.1556/112.2023.00186

Varga (2025): Varga, Livia, 2025, Forecasting cause-of-death mortality with single- and multi-population models in Hungary. *Scandinavian Actuarial Journal*, pp 1–32.
https://doi.org/10.1080/03461238.2025.2480846

Vékás (2017): Vékás, Péter, 2017, Nyugdíjcélú életjáradékok élettartam-kockázata az általánosított korcsoport-időszak-kohorsz modellkeretben. *Statisztikai Szemle* **95(2)**, pp 139–165.
https://doi.org/10.20311/stat2017.02.hu0139

Vékás (2018): Vékás, Péter, 2018, Változások a halandóságjavulás mintázatában Magyarországon. *Biztosítás és Kockázat* **5(3)**, pp 34–47.
https://doi.org/10.18530/BK.2018.3.34

Vékás (2019): Vékás, Péter, 2019, Az élettartam-kockázat modellezése. *Budapesti Corvinus Egyetem*, Budapest. ISBN 978-963-503-768-1
http://unipub.lib.uni-corvinus.hu/4112/1/elettartam_0612.pdf

Vékás (2020): Vékás, Péter, 2020, Rotation of the age pattern of mortality improvements in the European Union. *Central European Journal of Operations Research* **28**, pp 1031–1048.
https://doi.org/10.1007/s10100-019-00617-0

Villegas et al. (2017): Villegas, Andrés M. – Haberman, Steven – Kaishev, Vladimir K. – Millossovich, Pietro, 2017, A comparative study of two-population models for the assessment of basis risk in longevity hedges. *ASTIN Bulletin* **47(3)**, pp 631–679.
https://doi.org/10.1017/asb.2017.18

Villegas et al. (2018): Villegas, Andrés M. – Kaishev, Vladimir K. – Millossovich, Pietro, 2018, StMoMo: An R package for stochastic mortality modeling. *Journal of Statistical Software* **84(3)**, pp 1–38.
https://doi.org/10.18637/jss.v084.i03

Wen et al. (2021): Wen, Jie – Cairns, Andrew J. G. – Kleinow, Torsten, 2021, Fitting multi-population models to socio-economic groups. *Annals of Actuarial Science* **15(1)**, pp 144–172.
https://doi.org/10.1017/S1748499520000184

Wilmoth (1993): Wilmoth, John R., 1993, Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Technical Report, *University of California*, Berkeley, Department of Demography.
https://u.demog.berkeley.edu/~jrw/Papers/LCtech.pdf

Wilmoth et al. (2021): Wilmoth, John R. – Andreev, Kirill – Jdanov, Dmitri – Glei, Dana A. – Riffe, Timothy – Boe, Carl – Bubenheim, Michael – Philipov, Dimiter – Shkolnikov, Vladimir – Vachon, Pierre – Winant, Celeste – Barbieri, Magali, 2021, Methods protocol for the Human Mortality Database. *University of California*, Berkeley, and *Max Planck Institute for Demographic Research*, Rostock.
https://www.mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf

Wüthrich–Merz (2023): Wüthrich, Mario V. – Merz, Michael, 2023, Statistical foundations of actuarial learning and its applications. *Springer Actuarial*, Cham. ISBN 978-3-031-12408-2
https://doi.org/10.1007/978-3-031-12409-9

Yang et al. (2016): Yang, Bowen – Li, Jackie – Balasooriya, Uditha, 2016, Cohort extension of the Poisson common factor model for modelling both genders jointly. *Scandinavian Actuarial Journal* **2016(2)**, pp 93–112.
https://doi.org/10.1080/03461238.2014.908411

Zittersteyn–Alonso-García (2021): Zittersteyn, Geert – Alonso-García, Jennifer, 2021, Common factor cause-specific mortality model. *Risks* **9(12)**, pp 1–30.
https://doi.org/10.3390/risks9120221

Zivot–Andrews (1992): Zivot, Eric – Andrews, Donald W. K., 1992, Further evidence of the great crash, the oil-price shock and the unit-root hypothesis. *Journal of Business and Economic Statistics* **10(3)**, pp 251–270.
https://doi.org/10.1080/07350015.1992.10509904

Zonda et al. (2013): Zonda, Tamás – Paksi, Borbála – Veres, Előd, 2013, Az öngyilkosságok alakulása Magyarországon (1970–2010). *Központi Statisztikai Hivatal*, Budapest. ISBN 978-963-235-432-3
https://www.ksh.hu/docs/hun/xftp/idoszaki/pdf/muhelytanulmanyok2.pdf

Internet sources

Cairns, Andrew J. G., 2007, LifeMetrics open source R code for stochastic mortality modelling. (downloaded: August 2025)
https://www.macs.hw.ac.uk/~andrewc/lifemetrics/

Eurostat: Description of the Eurostat method for the calculation of the life expectancies at all ages. (downloaded: August 2025)
https://ec.europa.eu/eurostat/cache/metadata/Annexes/demo_mor_esms_an_1.pdf

Human Mortality Database. University of California, Berkeley, USA, and Max Planck Institute for Demographic Research, Germany. (downloaded: August 2025)
https://www.mortality.org/

Hungarian Central Statistical Office: BNO-X–BNO-IX correspondence table. https://www.ksh.hu/classification_bno_x (downloaded: August 2025)

Hyndman, Rob – Booth, Heather – Tickle, Leonie – Maindonald, John – Wood, Simon, 2023, demography: Forecasting mortality, fertility, migration and population data. R package version 2.0. (downloaded: August 2025)
https://CRAN.R-project.org/package=demography

R Core Team, 2021, R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
https://www.R-project.org/ (downloaded: August 2025)

Ševčíková, Hana – Li, Nan – Gerland, Patrick, 2022, MortCast: Estimation and projection of age-specific mortality rates. R package version 2.7-0. (downloaded: August 2025)
https://CRAN.R-project.org/package=MortCast

Trapletti, Adrian – Hornik, Kurt, 2024, tseries: Time series analysis and computational finance. R package version 0.10-58. (downloaded: August 2025)
https://CRAN.R-project.org/package=tseries

Zeileis, Achim – Hothorn, Torsten, 2002, Diagnostic checking in regression relationships. *R News* **2(3)**, pp 7–10. (downloaded: August 2025)
https://CRAN.R-project.org/doc/Rnews/ and https://CRAN.R-project.org/package=lmtest

Acknowledgments

I would like to thank Dávid Burka and László Kovács for their valuable comments and supervision. I am also grateful to the Cooperative Doctoral Programme of the Ministry of Culture and Innovation in Hungary financed by the National Research, Development and Innovation Fund and the Hungarian Central Statistical Office for the opportunity to analyze Hungarian data.

Online Appendices

https://github.com/LiviaVarga/Fitting-and-forecasting-multi-population-mortality-models-based-on-Hungarian-regional-data

https://github.com/LiviaVarga/Forecasting-cause-of-death-mortality-with-single-and-multi-population-models-in-Hungary

https://github.com/LiviaVarga/Rotation-of-the-age-varying-parameters-in-multi-population-mortality-models

https://github.com/LiviaVarga/Forecasting-multi-population-mortality-models-using-tree-based-machine-learning-methods

Publications of the author

Varga, Livia, 2020, Sztochasztikus népesség-előrejelzés magyar adatokon. Kézirat. Budapesti Corvinus Egyetem, Budapest.
https://unipub.lib.uni-corvinus.hu/7334/

Varga, Livia, 2021, A koronavírus-járvány miatti többlethalálozás becslése Magyarországon. Kézirat. Budapesti Corvinus Egyetem, Budapest.
https://unipub.lib.uni-corvinus.hu/7363/

Varga, Livia, 2023, Fitting and forecasting multi-population mortality models based on Hungarian regional data. *Regional Statistics* **13(5)**, pp 863–898.
https://doi.org/10.15196/RS130504

Varga, Livia, 2024, Rotation of the age-varying parameters in multi-population mortality models. *Scientia et Securitas* **4(4)**, pp 375–391.
https://doi.org/10.1556/112.2023.00186

Varga, Livia, 2025, Forecasting cause-of-death mortality with single- and multi-population models in Hungary. *Scandinavian Actuarial Journal*, pp 1–32.
https://doi.org/10.1080/03461238.2025.2480846