Stockholms universitet

Trends and Forecasts of Mortality in Sweden: A Comparative Analysis Using the Age-Period-Cohort and Lee-Carter Models

Natanael Blomberg

Masteruppsats i försäkringsmatematik
Master Thesis in Actuarial Mathematics

Masteruppsats 2025:7
Försäkringsmatematik
Juni 2025

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2025:7**
http://www.math.su.se

Trends and Forecasts of Mortality in Sweden: A Comparative Analysis Using the Age-Period-Cohort and Lee-Carter Models

Natanael Blomberg*

June 2025

Abstract

This thesis investigates long-term mortality trends in Sweden using two main demographic forecasting models: the Poisson Lee-Carter and Age-Period-Cohort (APC) models. Using mortality data and population exposure from 1751 to 2023, the models are suitable and evaluated in various age groups, focusing on the forecast of mortality rates from 2024 to 2050.

For model validation, we trained Lee-Carter between 1751 and 2000 and tested performance between 2001 and 2023 using multiple historical windows. For the final forecast 2024-2050, the model was re-estimated in the entire 1751-2023 dataset, assuming that mortality improvements follow a persistent trend over time. On the other hand, the APC model includes separate period and cohort effects, allowing more flexible generational patterns, and is trained on a shorter window to better capture structural changes.

Out-of-sample evaluations show that the Lee-Carter model performs better for mid-adult age groups when trained on a long historic window, and it achieves lower forecast errors in short-term projections for specific ages. In turn, the APC model provides better predictions for younger and older age groups when recent decades are highlighted, and it outperforms Lee-Carter over much longer forecast horizons. The forecast error is generally higher for the APC model due to the variability introduced by cohort effects. These results highlight a trade-off between the robustness and interpretability of the Lee-Carter model and the flexibility and responsiveness of the APC model to structural changes. In general, the findings support the use of both models as complementary tools in the analysis and prediction of mortality.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: natanaelblomberg@gmail.com. Supervisor: Mathias Millberg Lindholm.

Acknowledgments

I would like to take this opportunity to thank my supervisor Mathias Millberg Lindholm for his guidance and support during my master thesis.

AI Statement

This thesis has used AI tools for spell checking, grammar, and supporting idea development through discussions.

Contents

1 Introduction

Mortality trends and forecasts have significantly impacted demographic research, life insurance, and public policy planning. Mortality rates will never be fixed, so understanding and forecasting how they change over time will be of interest to healthcare resources and insurance pricing. Sweden is one of the best countries to collect mortality data. The mortality data used in this thesis date back to 1751, providing an unusually long historical record that enables the study of long-term survival trends and structural changes over nearly three centuries.

This thesis will analyze Swedish mortality using two well-known statistical models: the Lee-Carter and Age-Period-Cohort (APC) models. The reason for choosing these two models is because of how they describe mortality changes. Although both models aim to describe changes in mortality over time, they use different assumptions and structures, making them suitable for comparison. The Lee-Carter method was introduced in 1992 by Lee and Carter [6] and has since been used for long-term forecasts because it handles time index and age effects well. The APC model [9] has never really been discovered but instead developed over time. Similarly to Lee-Carter, APC also uses age and time but uses cohort as a third component to separate the effects differently. Using the Lee-Carter and APC model, the idea is to compare their performance when looking at the past and trying to forecast what happens next.

The data consists of mortality data for men and women between 0 and 110 years of age between 1751 and 2023. It is an extensive data set and gives a good opportunity to examine how mortality has changed over almost 300 years. The results will be studied separately for men and women and the population as a whole to see if there are any significant differences between them.

An out-of-sample test will validate how well the two models work with regard to forecasting. This will be done by selecting a train and test period, where we train the model over time and evaluate forecasts using data not used for training. This will help us determine how reliable the models are with regard to long-term forecasting.

The thesis is divided into five main parts. The first section is *Introduction* and summarizes the methods and main results of this thesis. Section 2, called *Theory*, explains the theory behind the Lee-Carter and APC models, including how they are built. The third section *Data*, goes through the data set and the steps taken to prepare

it. Section 4, *Analysis and Results*, shows the results with tables and graphs from the calculations and implementations of the theories. Finally, the *Discussion* section includes the conclusions, discussion about what worked well, and ideas for future enhancements.

The thesis compares the Lee-Carter and APC models using Swedish mortality data. Hopefully, the results can be helpful for those working with forecasts in insurance, public health, or demography.

2 Theory

2.1 Mortality Models

2.1.1 Survival Function

The survival function is defined as follows [1]:

The survival function represents an individual's probability of surviving beyond time x (i.e., experiencing the event after time x). The survival function is often defined as:

$$S(x) = P(X > x). \tag{1}$$

If X is a continuous random variable, then $S(x)$ is a continuous and strictly decreasing function.

When X is a continuous random variable, the survival function is the complement of the cumulative distribution function (CDF),

$$S(x) = 1 - F(x), \tag{2}$$

where $F(x) = P(X \leq x)$ is the CDF. Additionally, the survival function can be expressed as the integral of the probability density function (PDF), $f(x)$:

$$S(x) = \int_x^\infty f(t)\,\mathrm{d}t. \tag{3}$$

Taking the derivative of $S(x)$ gives the relationship between the survival function and the PDF:

$$f(x) = -\frac{\mathrm{d}S(x)}{\mathrm{d}\,x}. \tag{4}$$

Here, $f(x)\,\mathrm{d}\,x$ can be interpreted as the approximate probability that the event occurs at time x. Since $f(x)$ is a probability density function, it is nonnegative, and the total area under $f(x)$ integrates into one:

$$\int_0^\infty f(x)\,\mathrm{d}x = 1. \tag{5}$$

2.1.2 Hazard Function

The hazard function is defined as follows [1]:

An essential quantity fundamental in survival analysis is the hazard function. This function is also known as the force of mortality in demography, the intensity function in stochastic processes, the inverse of Mill's economic ratio, or simply the hazard function. The hazard function is defined by

$$h(x) = \lim_{\Delta x \to 0} \frac{P[x \le X < x + \Delta x \mid X \ge x]}{\Delta x}. \tag{6}$$

If X is a continuous random variable, then

$$h(x) = \frac{f(x)}{S(x)} = -\frac{\mathrm{d} \ln S(x)}{\mathrm{d}\, x}. \tag{7}$$

A related quantity is the cumulative hazard function $H(x)$, defined by

$$H(x) = \int_0^x h(u)\ \mathrm{d}\, u = -\ln S(x). \tag{8}$$

Thus, for continuous lifetimes,

$$S(x) = \exp[-H(x)] = \exp\left(-\int_0^x h(u)\ \mathrm{d}\, u\right). \tag{9}$$

From (6), it can be seen that $h(x)\Delta x$ can be viewed as the 'approximate' probability that an individual of age x experiences the event in the next instant. The hazard function helps determine appropriate failure distributions by describing how the risk of experiencing the event changes over time. There are many general shapes for the hazard rate. The only restriction is that $h(x)$ must be nonnegative, i.e., $h(x) \ge 0$, and we assume that $h(x)$ is continuous.

2.1.3 Cumulative Hazard Function and the Nelson-Aalen Estimator

Without parametric assumptions, the hazard rate $h(x)$ can be any nonnegative function, making direct estimation difficult. However, it is easier to estimate the cumulative hazard function [1].

$$H(x) = \int_0^x h(u)\,du, \tag{10}$$

without assuming any structure on $h(x)$. This is similar to estimating the cumulative distribution function. The result is the Nelson-Aalen estimator, given by:

$$\hat{H}(x) = \sum_{x_i \leq x} \frac{d_i}{Y(x_i)}, \tag{11}$$

where d_i is the number of events that occur at time x_i, and $Y(x_i)$ represents the number of individuals at risk just before time x_i. This estimator provides a nonparametric approximation of the cumulative hazard function based on observed event times and is widely used in survival analysis.

These fundamental concepts form the basis for analyzing mortality patterns. Forecasting techniques, such as the Lee-Carter and Age-Period-Cohort models, build on these ideas and are explored in subsequent sections [2].

2.2 Time Series Modelling

In forecasting mortality, time series models provide a framework to capture and predict patterns and dynamics in age-specific death rates. Incorporating time series techniques allows us to leverage the structure of past trends in order to generate more accurate and robust forward projections. One of the most well-known models is AutoRegressive Integrated Moving Average (ARIMA).

2.2.1 ARIMA Process

The ARIMA process is a versatile and powerful tool for univariate time series forecasting [3]. An ARIMA(p, d, q) model is defined by three nonnegative integers[4]:

- p: the order of the autoregressive (AR) component, representing how many past observations are used to model the current value.
- d: the degree of differencing, indicating how often the data have been differenced to achieve stationarity.
- q: the order of the moving average (MA) component, capturing the influence of past forecast errors on the current value.

In terms of y_t, the ARIMA(p, d, q) forecasting equation is

$$\Delta^d y_t = c + \phi_1 \Delta^d y_{t-1} + \cdots + \phi_p \Delta^d y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t.$$

where Δ is the difference, ϕ_i and θ_j are the AR and MA coefficients, c is the drift (i.e. constant), and ϵ_t is a Gaussian white noise process with variance σ^2 [5].

This thesis uses ARIMA$(0, 1, 0)$ with drift for the time-varying parameter k_t in Lee-Carter which gives

$$\Delta k_t = c + \varepsilon_t.$$

In APC ARIMA$(1, 1, 0)$ with drift will be used for γ_t and δ_c which gives

$$\Delta\delta_{t-x} \;=\; \phi\,\Delta\delta_{t-x-1} \;+\; c \;+\; \varepsilon_{t-x},$$

$$\Delta\gamma_t \;=\; \phi\,\Delta\gamma_{t-1} \;+\; c \;+\; \varepsilon_t.$$

The choice of these Arima(p, d, q) models will be discussed in Sections 2.3.4 and 2.4.3.

2.3 The Poisson Lee-Carter Model

The **Lee-Carter (LC) model**, introduced by Lee and Carter in their seminal paper [6], is one of the most widely used approaches to model and forecast mortality rates. Although initially developed for the U.S. population, it has been applied in numerous countries and subpopulations. However, it should be clarified that this thesis uses Poisson Lee-Carter.

2.3.1 Model Specification

Let $m_{x,t}$ denote the central death rate at age x in year t. The Lee-Carter model assumes a log-bilinear structure:

$$\log\big(m_{x,t}\big) \;=\; a_x \;+\; b_x\,k_t, \tag{12}$$

where

- a_x is an age-specific baseline capturing the overall level of mortality at age x,
- k_t is a calendar time effect capturing how mortality changes across all ages in year t,
- b_x represents how sensitive age x is to changes in k_t.

2.3.2 Parameter Estimation via SVD

To give a brief historical explanation of the Lee–Carter model, we first explain the original SVD-based fitting procedure.

Lee and Carter (1992) estimate (a_x, b_x, k_t) with a rank 1 Singular Value Decomposition (SVD) [6]. First set

$$a_x \;=\; \frac{1}{T}\sum_{t=1}^{T}\log(m_{x,t}), \qquad Z_{x,t} \;=\; \log(m_{x,t}) - a_x.$$

Applying SVD to Z gives the approximation

$$Z \;=\; U\,\Sigma\,V^{\top} = \sum_{i=1}^{r}\sigma_i u_i v_i^{T},$$

where $U_{(X\times X)}$ and $V_{(T\times T)}$ are orthogonal matrices and $\Sigma_{(X\times T)}$ is a rectangular diagonal matrix.

Retaining only the first component gives the rank 1 approximation $Z \approx \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\top}$. We set $b_x = u_{1,x}$ and $k_t = \sigma_1 v_{1,t}$, then rescale so that $\sum_x b_x = 1$ and $\sum_t k_t = 0$.

This is just a brief walk-through of the implementation; for the interested reader, see Lee and Carter (1992) [6].

2.3.3 Parameter Estimation via Poisson Maximum Likelihood

Following Brouhns, Denuit, and Vermunt [7], we model the death counts $D_{x,t}$ as

$$D_{x,t} \mid E_{x,t} \;\sim\; \text{Poisson}\big(E_{x,t}\, m_{x,t}\big),$$

where $E_{x,t}$ are the exposures and $m_{x,t} = \exp(a_x + b_x k_t)$. The parameters (a_x, b_x, k_t) are estimated by maximizing the Poisson log-likelihood under the identifiability constraints

$$\sum_x b_x = 1, \qquad \sum_t k_t = 0.$$

These constraints are needed because, without them, there is an infinite number of solutions for $\{a_x, b_x, k_t\}$ [8]. In practice, we obtain the MLEs via an iterative algorithm such as Newton–Raphson.

Once (a_x, b_x, k_t) have been estimated via maximum likelihood, a_x and b_x are held fixed across all ages, and future values of k_t are projected using an ARIMA process.

2.3.4 Forecasting the Time Index

We must project k_t for periods beyond the observed sample to predict future mortality. A common choice, originally advocated by Lee and Carter [6], is a random walk with drift:

$$k_{t+1} \;=\; k_t + c + \epsilon_{t+1}, \tag{13}$$

where c is a constant drift and ϵ_{t+1} is white noise.

In particular, this random walk with drift can be viewed as an ARIMA$(0, 1, 0)$ model [4]. Setting $p = q = 0$ is motivated by the structure of the Lee–Carter model, which assumes that mortality changes gradually over time without reverting to a long-term mean. The choice $d = 1$ captures this non-reverting trend by modeling k_t as a random walk with drift.

The key idea is that k_t captures the overall level of mortality improvement at all ages. Once the chosen ARIMA model is fitted, we forecast future values $\hat{k}_{t+h}$ for $h = 1, 2, \ldots$ years ahead.

Since a_x and b_x are fixed, the projected logarithmic mortality for year $(t + h)$ then becomes:

$$\log\big(\hat{m}_{x,\,t+h}\big) \;=\; \hat{a}_x \;+\; \hat{b}_x\,\hat{k}_{t+h}.$$

Exponentiating provides the forecast for mortality at each age x:

$$\hat{m}_{x,\,t+h} \;=\; \exp\Big(\hat{a}_x + \hat{b}_x\,\hat{k}_{t+h}\Big).$$

Thus, the future mortality of each age depends on the predicted trajectory of k_t. If $\hat{k}_{t+h}$ consistently decreases over time, it implies ongoing improvements in mortality; If it stabilizes or increases, the improvements slow or reverse. Using ARIMA methods, the model systematically captures any serial dependence (by regressing k_t on its own past values) as well as any drift in the historical k_t sequence, enabling more flexible forecasts.

2.4 Poisson Age Period Cohort (APC) Model

The Age-Period-Cohort (APC) model is widely used in demographic, epidemiological, and actuarial research to analyze mortality trends and distinguish the effects of age, period, and cohort. This section overviews the APC model, its identification challenges, estimation techniques, and applications. We would also like to specify that this thesis uses Poisson Age-Period-Cohort.

2.4.1 Components of the APC Model

The APC model decomposes the observed mortality rates into three distinct effects[9]:

- **Age Effect:** Represents aging-related biological and social changes. Mortality typically increases with age due to physiological deterioration and accumulated exposure to risk.
- **Period Effect:** Captures external events or societal changes that influence all age groups at a given time (e.g. pandemics, medical advancements, economic crises).
- **Cohort Effect:** Reflects generational influences, where individuals born in the same period may share common exposures, habits, or health risks (e.g. the prevalence of smoking and early life conditions).

The log mortality rate $m_{x,t}$ at age x and period t is expressed as:

$$\log(m_{x,t}) = \alpha_x + \gamma_t + \delta_c, \tag{14}$$

where:

- α_x represents the age effect,
- γ_t represents the period effect,
- δ_c represents the cohort effect, where $c = t - x$.

2.4.2 Parameter Estimation via Maximum Likelihood for APC

Following Section 2.3.3, the parameters $(\alpha_x, \gamma_t, \delta_c)$ are estimated by maximizing the Poisson log-likelihood subject to the identifiability constraints

$$\sum_t \gamma_t = 0, \quad \sum_c \delta_c = 0, \quad \sum_c c\,\delta_c = 0.$$

Similar to the Lee–Carter model, these constraints are imposed to avoid infinitely many equivalent solutions. In practice, the MLEs are obtained via an iterative process such as Newton–Raphson [10].

Once $(\alpha_x, \gamma_t, \delta_c)$ are estimated, α_x is held fixed while we can start projecting future γ_t and δ_c

2.4.3 Forecasting With The Age-Period-Cohort Model

To extrapolate mortality using the APC model, we follow the framework described by [11], where the logarithm of mortality is specified as in Equation (14). The model faces an identifiability issue because age, period, and cohort are linearly dependent. This is why the identifiability constraints are so important.

Once the period γ_t and cohort δ_{t-x} effects are estimated, we forecast them beyond the observed data using time-series methods that remove these identification-related linear trends.

As shown in Section 2.2.1, we use an ARIMA(1,1,0) model with drift to forecast the cohort effect δ_{t-x}. Concretely, we difference the estimated δ_{t-x} to obtain $\Delta\delta_{t-x} = \delta_{t-x} - \delta_{t-x-1}$, and then model this differenced series as an AR(1) process:

$$\Delta\delta_{t-x} \;=\; \phi\,\Delta\delta_{t-x-1} \;+\; c \;+\; \varepsilon_{t-x},$$

where ϕ is the autoregressive parameter, c is a drift term, and ε_{t-x} is white noise.

For the period effect γ_t, we apply a similar ARIMA$(1, 1, 0)$ specification. In the `StMoMo` software, this is sometimes labeled a "mean-reverting random walk with drift (mrwd)". Strictly speaking,

$$\Delta\gamma_t \;=\; \phi\,\Delta\gamma_{t-1} \;+\; c \;+\; \varepsilon_t$$

implies that the increments $\Delta\gamma_t$ partially revert toward zero if $|\phi| < 1$. This prevents long-run drift in γ_t without forcing γ_t itself to revert to a specific mean. Hence, "mean reversion" refers only to the differenced series, not a strict stationarity in γ_t. After forecasting $\Delta\gamma_t$ forward from the last observed year, we reintegrate to obtain predicted values of γ_t in the forecast horizon.

By combining these projected time-series paths for $\{\gamma_t\}$ and $\{\delta_{t-x}\}$ with the age effect $\{\alpha_x\}$, the APC model yields forecasts of log-mortality. Formally, for $t > T$ (future years),

$$m_{x,t} \;=\; \exp\!\big(\hat{\alpha}_x + \hat{\gamma}_t + \hat{\delta}_{t-x}\big),$$

where hats denote the estimated or forecasted parameter values. This approach accounts for separate period and cohort influences while respecting the identifiability constraints inherent in the APC framework.

2.5 Deviance and Residuals in a Poisson Framework

When fitting the Lee-Carter and APC models via a Poisson likelihood, deviance is the natural measure of lack of fit. We can write the Poisson likelihood according to

$$D = 2\sum_{x,t}\Big[y_{x,t}\ln\Big(\tfrac{y_{x,t}}{\hat{\mu}_{x,t}}\Big) - \big(y_{x,t} - \hat{\mu}_{x,t}\big)\Big], \tag{15}$$

where $y_{x,t}$ is the observed count of deaths $D_{x,t}$ at age x and time t, and $\hat{\mu}_{x,t}$ is the Poisson mean $E_{x,t}m_{x,t}$ fitted to the model (e.g.$\sim m_{x,t} = \exp(a_x + b_x k_t)$ for Lee-Carter) [12]. The corresponding deviance residual for each cell (x,t) is given by

$$r_{x,t} = \operatorname{sign}\big(y_{x,t} - \hat{\mu}_{x,t}\big)\sqrt{d_{x,t}}\,, \quad d_{x,t} = 2\Big[y_{x,t}\ln\Big(\tfrac{y_{x,t}}{\hat{\mu}_{x,t}}\Big) - \big(y_{x,t} - \hat{\mu}_{x,t}\big)\Big]. \tag{16}$$

The function $\operatorname{sign}(y_{x,t} - \hat{\mu}_{x,t})$ ensures that the residuals retain the direction of deviation. Specifically, it assigns a value of $+1$ when the observed death count $y_{x,t}$ exceeds the fitted value $\hat{\mu}_{x,t}$, indicating an underestimation by the model. In contrast, it assigns -1 when $y_{x,t} < \hat{\mu}_{x,t}$, indicating that the model has overestimated the observed death count. If $y_{x,t} = \hat{\mu}_{x,t}$, the residual is zero, representing a perfect fit.

$$\operatorname{sign}(z) = \begin{cases} 1, & \text{if } z > 0 \\ 0, & \text{if } z = 0 \\ -1, & \text{if } z < 0 \end{cases} \quad \text{where } z = y_{x,t} - \hat{\mu}_{x,t}. \tag{17}$$

This allows for an intuitive evaluation of whether the model systematically underestimates or overestimates mortality in different age-time groups.

Summing or averaging these residuals highlights age-time combinations where the model fits poorly. In this paper, we summarise goodness-of-fit by the Poisson deviance, to measure how well the Lee-Carter and Age-Period-Cohort model captures mortality patterns by age group.

2.6 Model Evaluation and Validation

We employ several validation metrics to assess the performance of the mortality models. The following subsections detail these evaluation methods.

2.6.1 Out-of-sample test (MAPE)

To evaluate the predictive accuracy of our models on data not used in parameter estimation, we employ an out-of-sample test using the Mean Absolute Percentage Error (MAPE). After fitting the models to the training period, we generate forecasts for the test period and compare them with the observed mortality rates. The MAPE is then calculated as follows [13]:

$$\text{MAPE} = \frac{100\%}{N} \sum_{t=1}^{N} \left| \frac{m_{x,t} - \hat{m}_{x,t}}{m_{x,t}} \right|, \tag{18}$$

where

- N is the number of observations in the test set,
- $m_{x,t}$ is the actual mortality rate in year t for age x, computed as $m_{x,t} = \frac{D_{x,t}}{E_{x,t}}$, where $D_{x,t}$ is the number of deaths and $E_{x,t}$ is the exposure for age x in year t,
- $\hat{m}_{x,t}$ is the predicted mortality rate for year t and age x.

A lower MAPE indicates better predictive performance, as the forecasts deviate less (in relative terms) from the actual observed rates. By examining the MAPE values for the Lee-Carter and Age-Period-Cohort models, we can directly compare their out-of-sample forecast accuracy. This measure is intuitive because it translates the average forecast error into a percentage, facilitating interpretation and comparison between different models or data sets.

2.6.2 Out-of-sample test (MAE)

To complement the percentage-based assessment provided by the MAPE, we also report the Mean Absolute Error (MAE), which measures the error between paired observations expressing the same phenomenon. After fitting the models to the training period, we generate forecasts for the test period and compute the MAE as

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^{N} \left| m_{x,t} - \hat{m}_{x,t} \right|, \tag{19}$$

where $N, m_{x,t}$ and $\hat{m}_{x,t}$ are described as in Section 2.6.1 [13].

A lower MAE indicates better predictive performance, as the forecasts deviate less (in absolute terms) from the observed rates. Unlike

MAPE, MAE retains the original units of the response variable, making it especially useful when percentage errors can be misleading. By comparing the MAE values across the Lee–Carter and Age–Period–Cohort models, we obtain a scale-sensitive view of out-of-sample accuracy that complements the relative perspective offered by the MAPE.

2.6.3 Out-of-sample test (RMSE)

To capture the effect of significant forecast errors, which might not work as well in absolute or percentage averages, we also report the root mean square error (RMSE). After estimating each model on the training period, we forecast the test period and compute the RMSE as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(m_{x,t} - \hat{m}_{x,t}\right)^2}, \tag{20}$$

where $N, m_{x,t}$, and $\hat{m}_{x,t}$ are defined as in Section 2.6.1 [13].

Because errors are squared before averaging, the RMSE assigns a greater weight to large deviations than the MAPE or the MAE. A lower RMSE, therefore, indicates better predictive performance, with a stronger penalty on forecasts that miss sharply. Like MAE, RMSE is expressed in the original units of the response variable, facilitating an intuitive interpretation of the magnitude of the forecast errors. By comparing RMSE values for the Lee–Carter and Age–Period–Cohort models, we gain an additional perspective on out-of-sample accuracy, highlighting models' sensitivity to occasional but severe forecasting misses.

2.6.4 Akaike information criterion

Akaike's Information Criterion (AIC) measures how well a model fits the data set without adding too many explanatory variables. When selecting the ARIMA(p, d, q) process to forecast k_t or (γ_t, δ_c), we use (AIC), assuming Gaussian errors for the time-series[14]:

$$\text{AIC} = -2\log L(\hat{\psi}) + 2\dim(\psi), \tag{21}$$

where $L(\hat{\psi})$ is the Gaussian likelihood function [15], and $\dim(\psi)$ is the number of free parameters for that ARIMA model.

This criterion is commonly used to compare different ARIMA(p, d, q) models, the one with the lowest score is the model favored by the AIC.

3 Data

To forecast Swedish mortality using the Lee-Carter and Age-Period-Cohort models, we require a data source that provides consistent, high-quality mortality measures over a long period. The Human Mortality Database (HMD) is an ideal choice, offering detailed records of deaths, exposures, and mortality rates for many countries, including Sweden [16]. We will use Swedish mortality data for the period 1751-2023 for people ages 0 to 110. The reason for the choice of this dataset is the amount of historical coverage. Since our data are of such a variety, we can explore different sets to analyze long-term trends and conduct robust forecasts.

The data obtained from the HMD website include various aspects, but the most important ones for us are the following.

- Number of deaths: The observed number of deaths occurred for a single age group during a specific calendar year, denoted $D_{x,t}$.
- Population exposures: An estimate of the number of people at risk during a year, illustrating how many people could experience mortality during that period, is denoted $E_{x,t}$.
- Central mortality rates: A ratio derived by the number of deaths divided by the corresponding exposures, denoted by $m_{x,t}$.

Since the HMD applies numerous quality checks to ensure consistency and reliability, only minimal data cleaning was necessary. We did not detect any missing or anomalous values, meaning that the data was already in a format appropriate to mortality modeling.

3.1 Stochastic Mortality Models in `StMoMo`

The `R` package used in this thesis is `StMoMo`, as it supports both the Lee–Carter and Age–Period–Cohort models. Implementing this package means that the theory behind how mortality rates are computed follows from Section 2, which discusses the Lee-Carter and Age-Period-Cohort models.

In practice, `StMoMo` takes the user-supplied mortality data $\{D_{x,t}, E_{x,t}\}$, the chosen model structure uses a generalized non-linear modeling approach to fit these parameters by maximum likelihood.

For our analysis we leverage these built-in facilities of `StMoMo` to calibrate both the Lee–Carter and APC models under the Poisson assumption,

$$D_{x,t} \,\big|\, E_{x,t} \;\sim\; \text{Poisson}\big(E_{x,t}\, m_{x,t}\big),$$

where $D_{x,t}$ is the observed number of deaths, $E_{x,t}$ the exposure, and $m_{x,t}$ the central death rate [10].

As noted above, the dataset spans 1751–2023 and we will select ages 20–90 in single-year increments. The reasoning behind the age span is to reduce the number of tables and figures. We calibrate each `StMoMo` model on three historical windows, 1751–2000, 1850–2000, and 1950–2000 and evaluate the forecasts on the period 2001–2023. Although Lee and Carter originally recommended training on a much shorter period, we include longer windows to examine how the choice affects the accuracy of the forecast. But more on this in Section 4.

4 Analysis and results

4.1 Exploratory Visualization

Figure 1 presents the historical mortality rates for Swedish men, women, and the combined population over 1751-2023. We observe a general decline in mortality across all ages, with a notably steeper decrease for younger and mid-life age groups. The male and female patterns visually appear similar, prompting a more detailed analysis below.

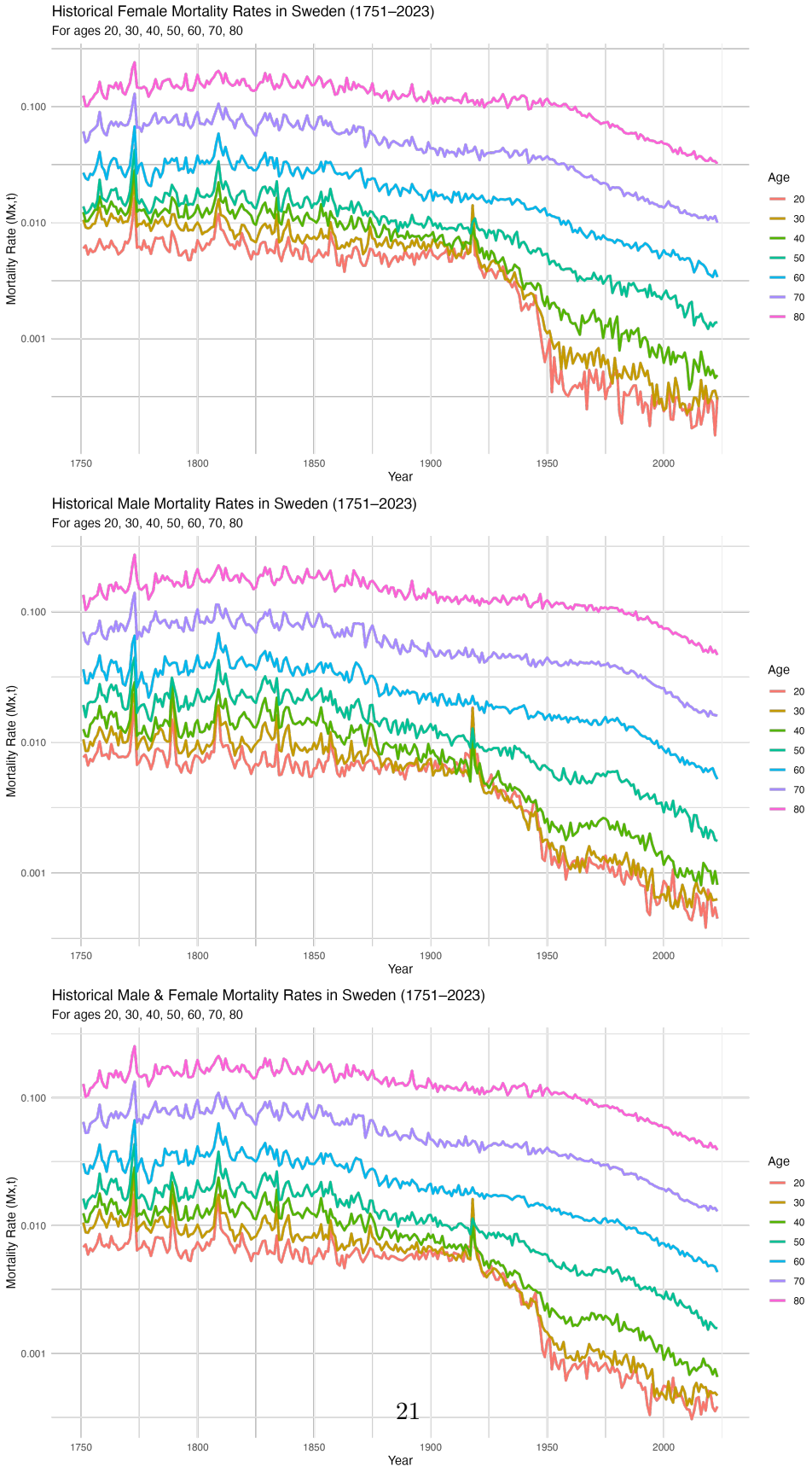

Figure 1: Historical Mortality Rates

4.2 Lee Carter method

In this section, we compare Swedish mortality data under the Lee-Carter model, examining separate fits for males vs. females and finally the combined death counts and exposures for both sexes. We also investigate different training windows and forecast mortality rates out to 2050. All analyses use mortality data for ages 20–90 unless otherwise specified.

The data for men and women look very much alike. Therefore, we want to keep them together for further analysis to obtain more data and trustworthy results. However, we can not make these assumptions from Figure 1. We will therefore look at the model parameters such as a, b, k, and Poisson deviance (PD) and the out-of-sample MAPE test to see if they are similar.

4.2.1 Lee-Carter Model: Male vs. Female Comparison

We apply the Poisson Lee-Carter model to Swedish female and male mortality data from 1751 to 2000, focusing on ages 20-90. After estimating the age-specific baseline parameters (a_x) and sensitivities (b_x), we evaluate goodness-of-fit with the Poisson deviance, computing an age-specific deviance PD by summing the deviance contributions across all years of the estimation window.

Because a table that ranges from ages 20 to 90 could be lengthy, we group ages into five-year intervals (20-24, 25-29, ..., 85-90) and then take the average of a_x, b_x, and PD_x within each group. A table of these metrics concisely summarizes how well the model fits across broader age groups.

Table 1 presents these grouped results. The a column reflects the baseline level of log-mortality for each group, while b captures how sensitive that group's mortality is to the overall time trend $\{k_t\}$. Table 2 shows the mean of the PD values, indicating which age ranges tend to deviate the most under the Poisson-Lee-Carter model.

Table 1: Grouped summary of Lee-Carter parameters for Swedish males and females by five-year age groups (1751-2000).

AgeGroup	$\overline{a_x}$ **(F)**	$\overline{a_x}$ **(M)**	$\overline{b_x}$ **(F)**	$\overline{b_x}$ **(M)**
20-24	-5.59	-5.27	0.0197	0.0196
25-29	-5.48	-5.25	0.0208	0.0216
30-34	-5.35	-5.17	0.0216	0.0224
35-39	-5.21	-5.02	0.0203	0.0216
40-44	-5.01	-4.78	0.0193	0.0207
45-49	-4.82	-4.52	0.0160	0.0183
50-54	-4.52	-4.20	0.0146	0.0160
55-59	-4.20	-3.89	0.0129	0.0130
60-64	-3.78	-3.50	0.0124	0.0112
65-69	-3.33	-3.10	0.0111	0.00934
70-74	-2.84	-2.66	0.0100	0.00828
75-79	-2.36	-2.21	0.00844	0.00699
80-84	-1.90	-1.77	0.00669	0.00569
85-90	-1.42	-1.30	0.00507	0.00451

Table 2: In-sample grouped summary of Lee-Carter Poisson deviance for Swedish males and females by five-year age groups (1751-2000).

AgeGroup	$\overline{PD}$ **(F)**	$\overline{PD}$**(M)**
20-24	543.31	705.46
25-29	447.51	447.03
30-34	273.49	218.98
35-39	176.07	126.60
40-44	128.90	95.04
45-49	92.09	83.32
50-54	134.53	134.05
55-59	154.61	171.76
60-64	224.48	277.24
65-69	231.15	323.75
70-74	274.83	327.98
75-79	304.92	245.03
80-84	262.70	166.85
85-90	186.16	124.07

As we can see from the table above, the in-sample PD look very much alike. In some age groups the female model achieves a slightly smaller PD, whereas in others the male model fits marginally better. We view as positive since this gives us more information about that specific age group. Also, no one is substantially better than the other

across the board, implying that a suitable idea would be to combine these.

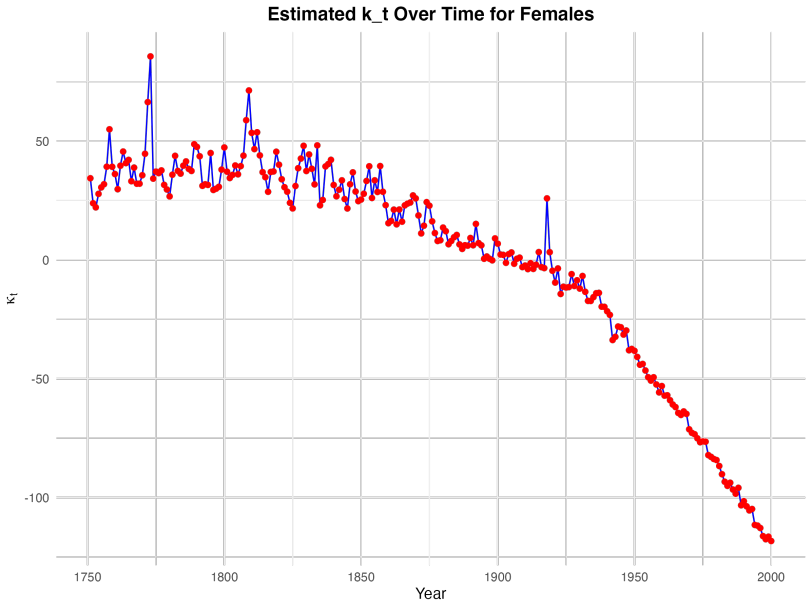

Figure 2: Female k_t values from 1751-2000 with ages 20-90

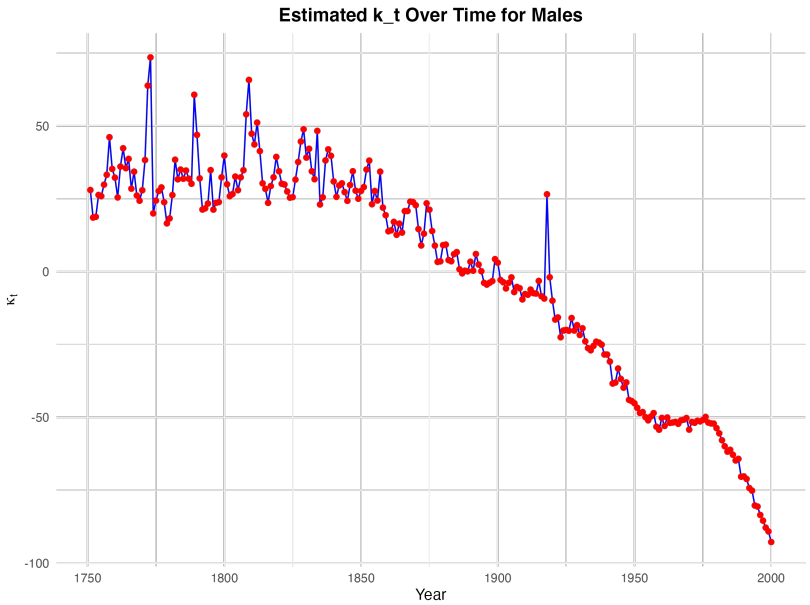

Figure 3: Male k_t values from 1751-2000 with ages 20-90

The k_t values in Figures 2 and 3 illustrate that the overall mortality

trend decreases over time. This observation is expected, primarily due to improvements in medical care and increasing medical knowledge. Nevertheless, we observe certain outliers, especially around the year 1919. A key contributing factor to this anomaly is the Spanish flu, which occurred during 1918-1919 and claimed roughly 39 000 Swedish lives [18]. It may be a good idea to remove these outliers and investigate how well our model fits the data once such irregularities occur, but more on this under Section 4.2.5.

The average a_x is most negative at younger ages and steadily becomes less negative (i.e., larger on the log scale) with increasing age group. These results align with the expectation that baseline mortality rates rise as we move from young adulthood to older ages.

The mean b_x is somewhat higher in early adulthood (around 0.02) and gradually declines for older age groups, indicating that younger adult mortality is slightly more sensitive to the overall mortality index k_t.

Younger adults (20-24) and some older groups (70-79) exhibit higher average PD, suggesting that the Poisson Lee-Carter model may fit those age segments less tightly or that those segments have more volatile mortality counts over time. Meanwhile, mid-adults (45-49) have relatively small PD, implying a closer fit in that range.

Overall, these findings are consistent with typical patterns in mortality modeling: the log-bilinear structure captures mid-adult ages quite well, whereas very young and older ages often introduce more significant variability.

To evaluate how well the Poisson Lee-Carter model performs in predicting future mortality, we conduct an out-of-sample test for 2001-2023. The model is fitted to Swedish female and male mortality data (ages 20-90) for 1751-2000; then forecasts are generated for 2001-2023. We compare these forecasts against the actual observed rates during that interval and measure accuracy using the Mean Absolute Percentage Error (MAPE), as defined in Equation (19).

Since MAPE can vary substantially by age (particularly for low mortality at younger ages), we further group the results into five-year age intervals and compute the average MAPE in each interval. Table 3 summarizes the findings: the highest MAPE appears at younger adult ages (20-24), reflecting the difficulty of accurately forecasting very low mortality rates. MAPE generally improves through midlife, reaching a

low of about 6% in ages 65-69, before rising again at older ages, likely due to greater mortality volatility in later life.

Table 3: Mean Absolute Percentage Error (MAPE) for Swedish males and females by five-year age groups (2001-2023).

AgeGroup	**MAPE (F)**	**MAPE (M)**
20-24	43.35	24.19
25-29	28.96	14.65
30-34	18.50	16.37
35-39	15.75	13.01
40-44	17.38	11.95
45-49	14.95	12.56
50-54	14.98	13.24
55-59	13.08	15.86
60-64	11.76	19.67
65-69	6.42	27.79
70-74	12.28	33.89
75-79	24.70	36.56
80-84	26.82	29.87
85-90	17.36	16.44

The results confirm that the Poisson Lee-Carter model forecasts most accurately in the mid-adult range, where mortality is neither extremely low nor highly volatile. Higher relative errors typically arise at the youngest ages because mortality rates are so low that even minor absolute discrepancies inflate percentage errors. Older ages can also show higher observed MAPE due to greater volatility in late-life mortality and the simplifying single-factor structure. It is worth noting that, historically, relatively few individuals survived to very old ages, so the available data can be sparse. This sparsity may lead to overfitting a handful of observations and thus producing an artificially low MAPE in certain older age intervals.

The MAPE values also confirm that the male and female data are not too far apart. Thus, we continue our analysis with a combined female and male data set. More analysis on male and female data can be seen in Section 6.

4.2.2 Combined data

Now that we have combined the Female and Male dataset, we can start by looking at the parameters below.

Table 4: Grouped summary of Lee-Carter parameters and Poisson deviance for Swedish males and females combined by five-year age groups (1751-2000).

AgeGroup	$\overline{a_x}$	$\overline{b_x}$	$\overline{PD}$
20-24	-5.41	0.0195	631.67
25-29	-5.35	0.0211	459.24
30-34	-5.25	0.0219	242.67
35-39	-5.11	0.0209	142.62
40-44	-4.88	0.0199	102.33
45-49	-4.66	0.0170	76.06
50-54	-4.35	0.0152	127.94
55-59	-4.04	0.0128	155.68
60-64	-3.64	0.0116	250.30
65-69	-3.22	0.0101	271.26
70-74	-2.75	0.0092	290.69
75-79	-2.29	0.0079	273.68
80-84	-1.84	0.0065	235.34
85-90	-1.37	0.0052	177.10

The results presented in Table 4 show the estimated Lee-Carter parameters for different five-year age groups based on combined mortality data for Swedish males and females between 1751 and 2000.

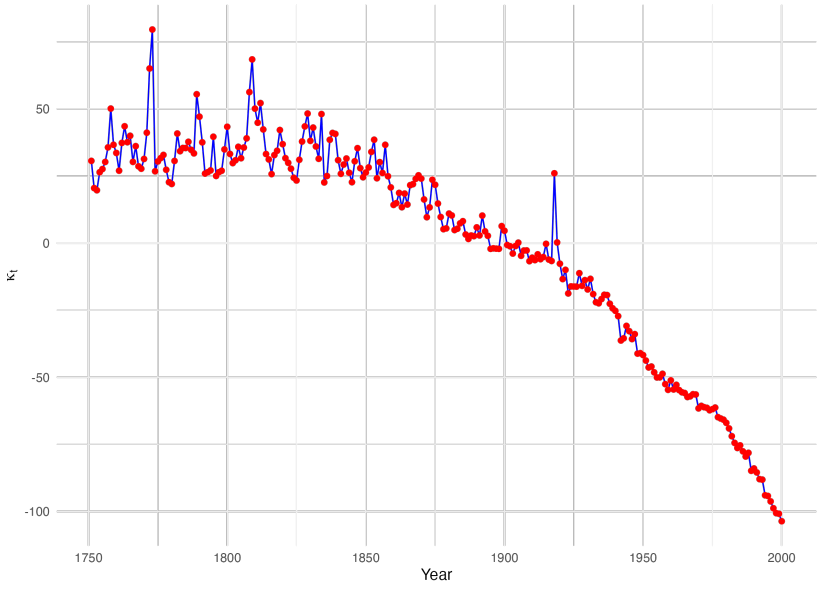

Figure 4: Female and male combined k_t values from 1751-2000 with ages 20-90

Figure 4 presents the estimated k_t values for the combined female

and male data from 1751-2000. As in the separate analyses, we observe a general downward trend, reflecting the long-term decline in mortality risks across these age groups. However, some notable spikes appear, most prominently around 1919. These outliers are again consistent with the observed effects of the Spanish flu during 1918-1919, which caused a sharp, temporary increase in mortality.

By combining both populations, we capture a more comprehensive picture of the overall mortality trajectory. Even though analyzing men and women individually may offer valuable insights into gender-specific health and longevity trends. Identifying and possibly mitigating the impact of outliers (especially those resulting from significant events such as pandemics) remains an important step in producing reliable mortality estimates and models.

The parameter a_x represents the baseline log-mortality for each age group. The values are more negative for younger age groups and increase with age, reflecting the well-established demographic pattern that mortality rates are lower at younger ages and rise with increasing age. The steepest increase in a_x occurs in mid-adulthood, which aligns with known mortality trends.

The b_x parameter captures the rate at which mortality changes over time for each age group. Higher b_x values indicate that mortality for that age group is more sensitive to period effects (such as medical advancements or economic fluctuations). The results show that b_x is highest for young adults (ages 25–39) and then declines gradually for older age groups. The decline makes sense to us since mortality at younger ages is more responsive to improvements in healthcare while aging processes and long-term trends influence mortality at older ages.

The Poisson deviance (PD) quantifies the fit of the Lee-Carter model across different age groups. Higher PD values suggest the model has more difficulty capturing mortality patterns for specific age groups. The largest PD values are visible for young adults (20–24 and 25–29), possibly due to the high variability in mortality at younger ages (e.g. accidents, infectious diseases). The older age groups (70+) also have relatively high PD values, likely due to more significant uncertainty in late-life mortality trends.

The patterns observed in a_x, b_x, and PD are consistent with previous mortality studies. Mortality rates tend to be more stable in the

middle ages, explaining why the PD are lowest between 40–60 years. Meanwhile, higher PD values can be seen at young and old age which indicates that those age groups experience more mortality volatility and period effects, making them harder to predict accurately.

These results might mean that forecasting future mortality trends will be most reliable for middle-aged groups (40–60 years). In contrast, special care must be taken when interpreting forecasts for young adults and the elderly, as these groups exhibit higher uncertainty.

We evaluated the prediction ability of the Poisson Lee-Carter model using three training windows: 1751-2000, 1850-2000, and 1950-2000. The reasoning behind these windows is to see which model performs best based on different training. After fitting the model in each case, we projected mortality for the same age spans as earlier, from 2001 through 2023. Table 5 presents the mean absolute percentage error (MAPE) between predicted and observed mortality rates in these out-of-sample years, with lower values indicating closer alignment with reality.

For each training window we first estimate the Poisson Lee-Carter parameters $\{a_x,\, b_x,\, k_t\}$ using maximum likelihood. The age-specific terms a_x and b_x are then held fixed, while the period index k_t is extrapolated to 2023 using an ARIMA$(0, 1, 0)$ with drift model.

The forecasts of age-specific central mortality rates are then obtained as

$$\hat{m}_{x,t} \;=\; \exp\big(a_x + b_x\, \hat{k}_t\big), \qquad t = 2001, \ldots, 2023, \tag{22}$$

and compared with the observed rates $m_{x,t}$ from the Human Mortality Database. We then use equation (18) to obtain the MAPE values below.

Table 5: Out-of-sample forecast accuracy of the Poisson Lee-Carter model: Mean Absolute Percentage Error (MAPE,%) for ages 20-90 when the model is trained on three historical windows (1751-2000, 1850-2000, and 1950-2000) and projected to 2001-2023.

	MAPE (%)		
AgeGroup	**1751-2000**	**1850-2000**	**1950-2000**
20-24	20.07	38.15	16.13
25-29	12.32	43.98	14.63
30-34	14.54	39.28	16.57
35-39	10.27	31.53	26.63
40-44	11.75	22.64	31.41
45-49	11.46	11.76	28.63
50-54	11.77	9.74	21.50
55-59	10.37	9.46	17.65
60-64	9.13	12.58	13.97
65-69	15.33	18.18	10.39
70-74	22.59	24.98	7.86
75-79	29.04	27.16	6.08
80-84	26.31	21.43	2.80
85-90	15.20	9.47	3.87

These results highlight how the choice of training period influences the accuracy of the forecast for 2001-2023. In Lee and Carters original paper, [19] they noted that forecasts were pretty stable when the training window was between 30 and 90 years but that shorter windows (10-20 years) introduced instability. We therefore investigated this in Table 11, Section 6. From Table 11 we see that we did not get better results when following Lee-Carter's recommendation, which is why we went for the spans above. In the following, we consider three historical spans to see which yields the best performance across age groups.

- 1751-2000. The longer historical window works best for younger and middle-aged groups (20-69). Including data over two centuries seems to smooth out anomalies and provide robust estimates for these ages. However, events such as the 1918-1919 Spanish Flu, which heavily impacted certain age groups, are also incorporated into this long history [18].

- 1850-2000. This intermediate span performs the worst for ages 20-39, indicating higher volatility in mid-1800s mortality data. The introduction of new medical interventions and changes in public health during that period may have generated complex trends that the model struggles to generalize for these age brackets.

- 1950-2000. Although [19] argued that basing the model on more recent decades can sharpen forecast accuracy, our results show that using only 1950-2000 data improves predictions for older ages (80+) but degrades them at younger ages (20-34). These results reflect the stronger impact of modern medical advancements, such as the 1958 invention of the pacemaker on older populations [20], along with the reduced relevance of historical epidemics (like the Spanish Flu) for contemporary elderly cohorts.

- **Younger ages (20-39).** These groups consistently exhibit higher error rates, regardless of the training period. These findings align with the unpredictability of mortality at young adult ages, which accidents, unknown epidemics, or shifts in lifestyle can strongly influence.

- **Middle-aged (40-69).** Forecast errors remain relatively low, especially when using the longer historical context (1751-2000). This suggests that mortality patterns in midlife are comparatively stable over time.

- **Older ages (70-90).** The most recent window (1950-2000) generally yields better forecasts for these cohorts, presumably because modern longevity improvements (e.g. post-1950 medical technology and increased hospital access) dominate late-life mortality trends. In contrast, older historical data may dilute the model with patterns less relevant to contemporary older-age mortality.

These results confirm that the best training period depends on the specific age group of interest. Longer spans may help younger and middle-aged cohorts by smoothing out temporary events. At the same time, more recent data can better capture ongoing progress in elderly care and life-extending medical innovations.

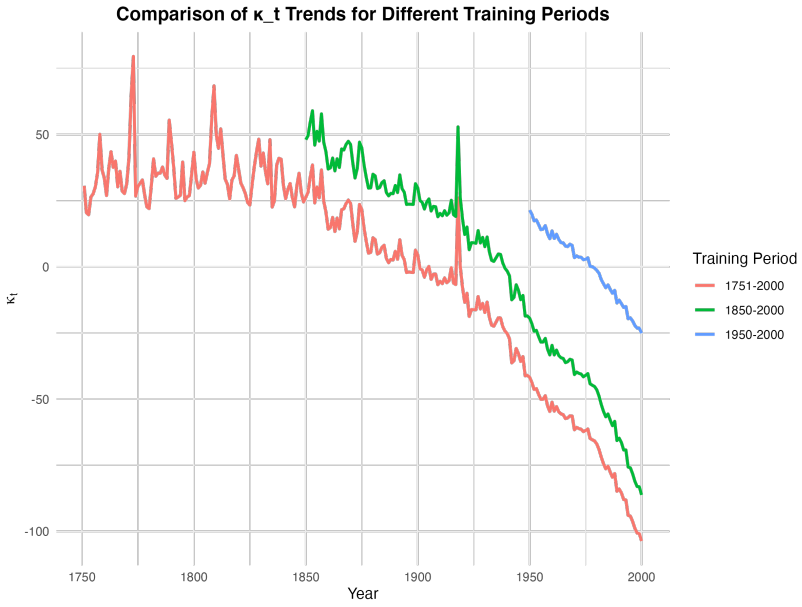

Figure 5: Different k_t values based on different training sets

Figure 5 compares the estimated k_t values obtained by fitting the mortality model on three historical spans: 1751-2000 (red), 1850-2000 (green), and 1950-2000 (blue). All three series show the same downward trend, but their levels and slopes differ because k_t is jointly identified with a_x and b_x. Changing the training window therefore changes a_x and b_x, which in turn rescales k_t.

Training over the most extended interval (1751-2000) incorporates a wide range of mortality behaviors and events (e.g. mid-19th century epidemics), producing a more elevated k in earlier decades. The medium-range interval (1850-2000) excludes some of the oldest data but still captures major shifts in mortality around the turn of the 20th century. Finally, the shorter and more recent interval (1950-2000) naturally places a stronger emphasis on modern mortality patterns, resulting in lower overall k levels prior to 1950 and a tighter alignment with contemporary trends in the second half of the century.

Each series still exhibits a visible spike around 1919, which aligns with the mortality impact of the Spanish Flu pandemic. However, the amplitude can differ depending on whether that event is included in the fitting window. Ultimately, these variations highlight the sensitivity of the Lee-Carter model to the training span: longer periods yield a broader historical context at the cost of diluting recent trends, while shorter windows capture current mortality patterns more closely but may lose valuable historical perspective.

Out-of-sample testing in 2001-2023 showed that the model estimated in the longest historical span (1751-2000) produces the lowest overall MAPE and shows better stability (Table 5). We therefore adopt this specification for our long-term projections. To exploit all available information, we now re-estimate the Poisson Lee-Carter model on the full period 1751-2023 and use it to forecast mortality rates for 2024-2050.

4.2.3 Forecasting Future Mortality Rates (2024–2050)

To forecast mortality rates for 2024–2050, we apply the Lee-Carter model following the methodology outlined in Section 2.3. Specifically, the time index k_t is modeled as a stochastic process and forecasted using a time series approach. For a detailed derivation and discussion of the Lee-Carter model, see Section 2.3. Below, we outline the specific forecasting steps and implementation details.

1. Fit a stochastic ARIMA$(0, 1, 0)$ process (random walk with drift), $k_{t+1} = k_t + c + \epsilon_{t+1}$, to the k_t series for 1751-2023.
2. Forecast k_t for 2024–2050 while the a_x and b_x estimates are held fixed.
3. Compute the predicted log-mortality rates for each age x in years $t = 2024, \ldots, 2050$:

$$\log \hat{m}_{x,t} = \hat{a}_x + \hat{b}_x \hat{k}_t. \tag{23}$$

4. Exponentiate to obtain the predicted mortality rates:

$$\hat{m}_{x,t} = \exp(\hat{a}_x + \hat{b}_x \hat{k}_t). \tag{24}$$

We are forecasting every age individually but want keep the age spans from earlier. We therefore compute the average mortality rate for the ages in that span.

The forecasted mortality rates $\hat{m}_{x,t}$ for 2024–2050 provide an estimate of future mortality trends. These estimates will be analyzed and compared with those from other models, such as the Age-Period-Cohort (APC) model, to assess their accuracy and predictive power [19].

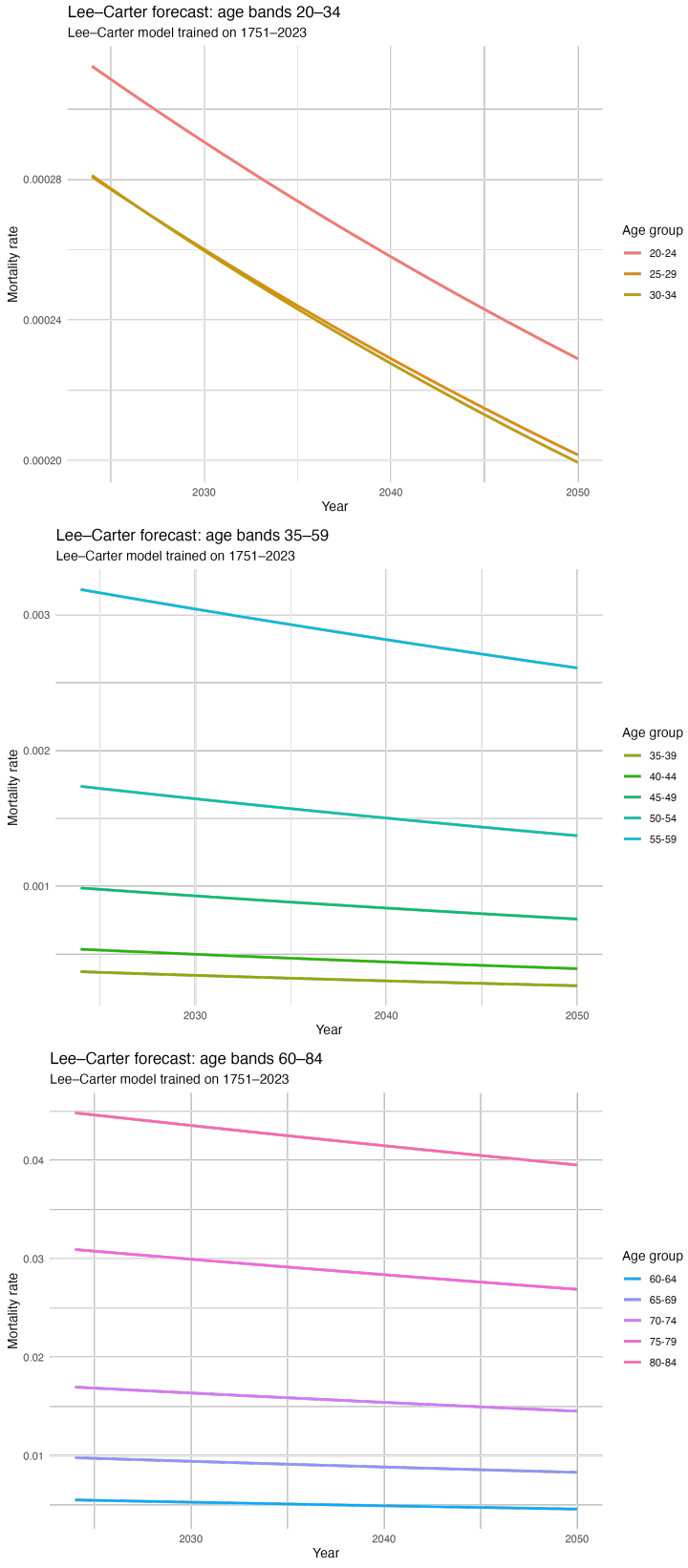

Figure 6: Forecasted Mortality Rates (2024–2050) using the Lee-Carter Model. The analysis is conducted for five-year age groups, ignoring the last group, 85-90, to avoid too many graphs.

Figure 6 show a steady decline in projected mortality across all ages, continuing the historical trend of improving longevity.

A closer look at Figure 6 reveals that the three lowest spans (20–24, 25–29, 30–34) almost overlap and decline heavily. Young-adult mortality has decreased considerably over the past 250+ years, so the estimation assigns the largest b_x coefficients to the youngest ages. When the period index k_t continues its downward drift, the product $b_x\, k_t$ drives a proportionally larger reduction in $\log(m_{x,t})$ for ages with large $|b_x|$, hence the steeper decline for ages 20–34.

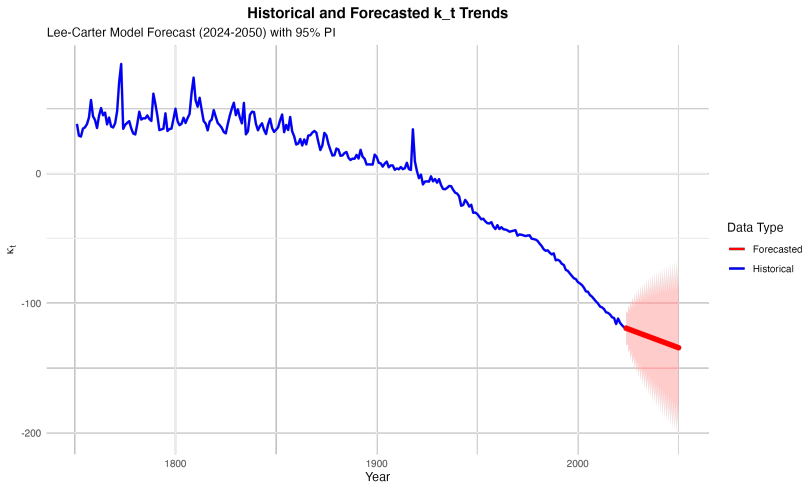

Figure 7: Forecasted k values from 2024-2050 with 95% Prediction Interval

Figure 7 shows the forecasted k_t values from 2024 to 2050, along with the corresponding prediction interval (highlighted in red), that is given via the `StMoMo` package. The model is calibrated on historical data from 1751 to 2023 and then extrapolated forward to predict the future mortality trend. The downward trajectory of k implies a continued decrease in mortality over the forecast horizon, consistent with longer-term improvements observed in the historical data.

The width of the prediction interval reflects the stochastic variation generated by the time-series model for k_t. While the central forecast suggests a relatively smooth continuation of current trends, the prediction span reminds us that actual outcomes may deviate from the predicted path due to unforeseen events, shifts in healthcare practices, or demographic changes.

4.2.4 Parametric Bootstrap Under Reduced Exposure

We conducted a parametric bootstrap to investigate how the Lee-Carter model behaves when exposure counts $E_{x,t}$ are systematically scaled down. We started by fitting the standard Lee-Carter model.

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

to obtain baseline estimates $\hat{a}_x$, $\hat{b}_x$, and $\hat{k}_t$. Then we applied a scale factor s to the original exposures, producing $\hat{E}_{x,t} = sE_{x,t}$. To incorporate the inherent Poisson variance of mortality data, we sample new deaths from a Poisson distribution with mean $\mu_{x,t} = \hat{m}_{x,t}\,\hat{E}_{x,t}$, where $\hat{m}_{x,t}$ are the originally fitted rates. Because

$$\mathrm{Var}(D_{x,t}) \;=\; \mu_{x,t}, \quad \text{and} \quad \hat{m}_{x,t} \;=\; \frac{D_{x,t}}{E_{x,t}},$$

this simulation step inherently captures the random fluctuations of $\frac{D_{x,t}}{E_{x,t}}$. Finally, we re-fitted the Poisson Lee-Carter model to each downscaled dataset, repeating this procedure 50 times for several s values.

In practice, mortality data may be incomplete or sparse. By artificially reducing exposure and resampling deaths under the Poisson assumption, we can see when Lee-Carter parameter estimates $(\hat{a}_x, \hat{b}_x, \hat{k}_t)$ become too inconsistent to be reliable. If s is moderately large (e.g. 0.1), the re-fitted parameters remain near their original values, implying that the model tolerates a 90% reduction in exposure. However, for very small s (e.g. 0.01 or 0.005), the estimates deviate sharply, indicating that the model breaks once the data becomes too sparse to capture age-specific trends accurately. Reducing our data this much makes the expected deaths in many age-year cells fall below one. The Poisson noise overwhelms the age pattern that the Lee–Carter model is trying to capture.

Figure 8 shows how the time-varying parameter k_t evolves under different scale factors. At $s = 0.1$, the bootstrapped trajectories stay relatively close to the baseline fit, reflecting stable estimation. As s decreases, these lines become more erratic and diverge from the original, illustrating the rapid loss of precision of the model under insufficient data.

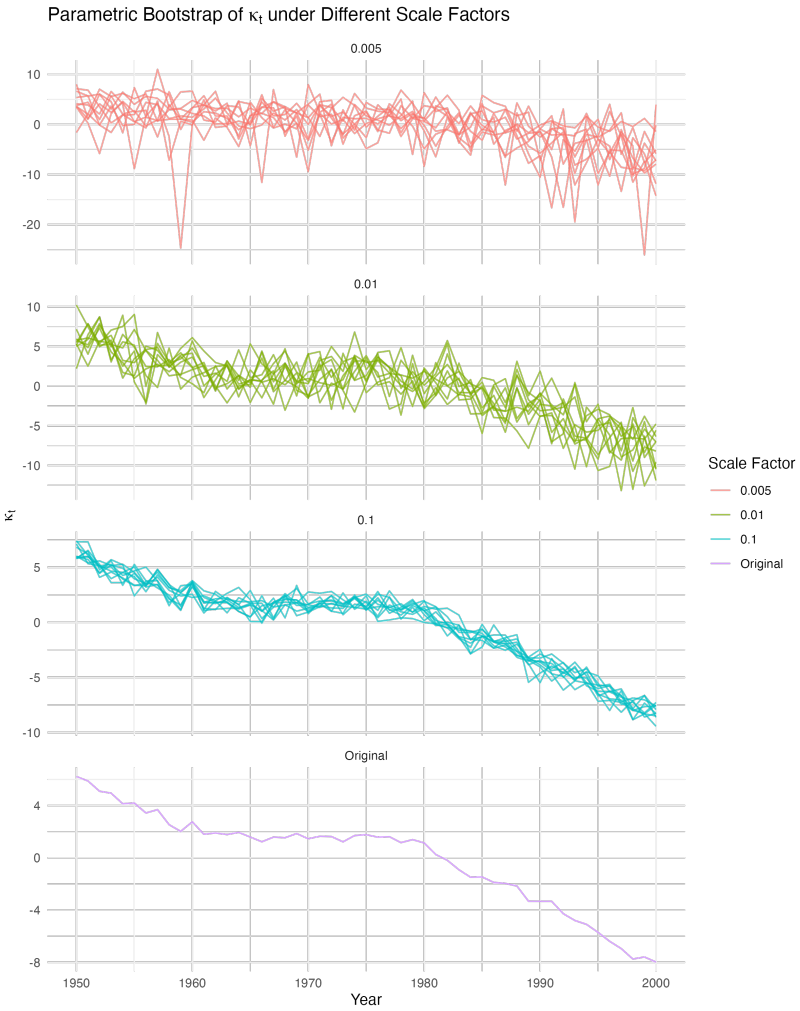

Figure 8: Example of bootstrapped k_t (ages 40-60, 1950-2000) at different scale factors. Smaller s increases variability in k_t, reflecting model breakdown.

Due to computational constraints, we focus on ages 40-60. A broader age range might reveal instability at larger s values or further nuance in the breakdown point, but the underlying approach remains the same. By combining Poisson-based resampling with scaled exposure, we directly assess both the mortality rate variance and the effect of reduced data quantity, illustrating when Lee-Carter fails to provide stable estimates of k.

4.2.5 Removing outliers

Future work may benefit from addressing data quality by detecting and removing outliers, particularly in older historical data or age groups with sparse observations. This could reduce the model's sensitivity and improve the forecast's accuracy. However, we have chosen to keep all observations. Rare spikes in mortality often reflect genuine events of demographic or historical interest, and the boundary between a true anomaly and an extreme value can be hard to tell.

4.3 Age Period Cohort

The next step in our study involves an analysis of Age–Period–Cohort (APC), a well-established framework that disentangles the separate contributions of **age**, **period** and **cohort** effects on mortality. Unlike the Lee–Carter model, which captures a single period index k_t modifying an age-specific baseline a_x, the APC model explicitly models age, period, and cohort effects separately (Section 2.4).

This procedure yields both forecasts and, by simulating the future paths, prediction intervals that reflect the uncertainty in extrapolating the period and cohort effects.

This section will use the combined male and female dataset to ensure our analysis aligns well with the evaluation using Lee-Carter.

4.3.1 Model Validation

Table 6: APC Summaries by 5-Year Age Groups. "Mean α" refers to the average age effect α_x across all ages in the bin, and "Mean PD" is the Poisson deviance for those ages.

AgeGroup	$\overline{\alpha_x}$	$\overline{PD}$
20-24	-5.62	765
25-29	-5.54	433
30-34	-5.43	245
35-39	-5.31	198
40-44	-5.10	212
45-49	-4.91	133
50-54	-4.61	151
55-59	-4.30	155
60-64	-3.90	187
65-69	-3.47	151
70-74	-3.00	114
75-79	-2.52	126
80-84	-2.04	237
85-90	-1.54	263

Table 6 summarizes two key indicators from an Age-Period-Cohort (APC) model fit on Swedish mortality data. The first is the average estimated age effect, α_x, across each 5-year age bin. Lower (more negative) values typically indicate lower baseline mortality levels than higher (less negative) values. The Poisson deviance in that age range reflects how well the APC model's Poisson assumption and estimated rates capture the observed death counts. Smaller sums indicate a better fit (less discrepancy), whereas more considerable sums can signal a poorer fit for those ages.

We group ages in 5-year intervals (e.g. 20-24, 25-29) to provide a concise, high-level comparison instead of listing individual ages. This allows us to identify broad patterns, such as consistently higher or lower PD values in particular segments of the lifespan.

The α_x term in the APC model corresponds to a long-term baseline mortality pattern by age after accounting for period and cohort effects. Because both the APC and Lee-Carter models (in `StMoMo`) are fitted using the same Poisson likelihood, their deviance are directly comparable. This allows a fair assessment of model fit across age, period, and cohort dimensions.

Thus, the table helps highlight how mortality behaves differently

across age groups and how well the model explains variation in deaths at each life stage.

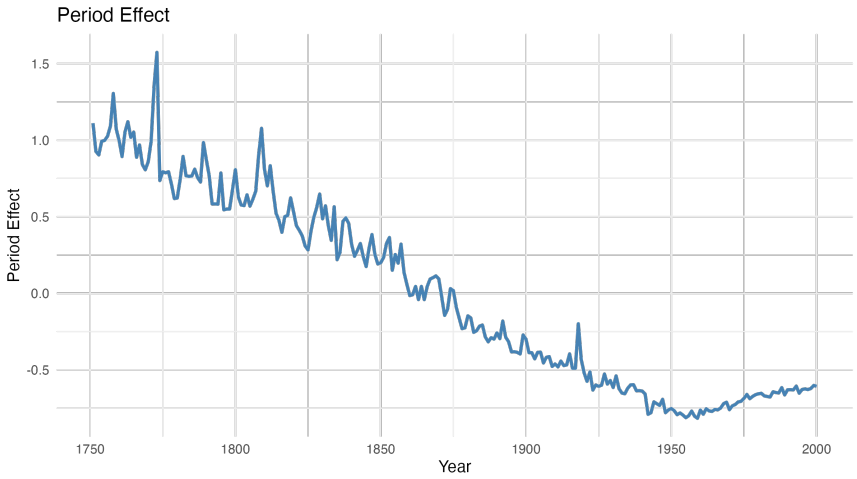

Figure 9: Estimated period effects ($\hat{\gamma}_t^f$) from the APC model. These effects capture temporal changes in mortality that affect all ages simultaneously.

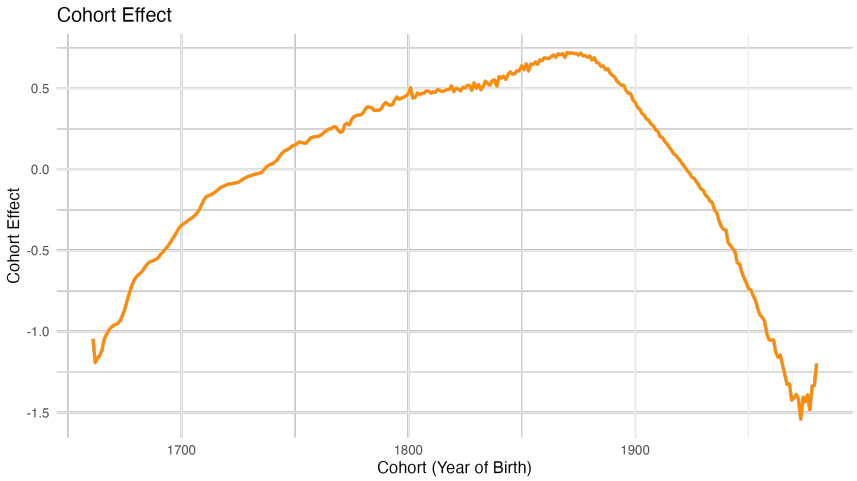

Figure 10: Estimated cohort effects ($\hat{\delta}_{t-x}^f$) from the APC model. These effects represent systematic differences in mortality across birth cohorts.

Figure 9 and 10 visualize the remaining components of the APC model: the period and cohort effects. The estimated period effect, $\hat{\gamma}_t^f$, in Figure 9, reflects broad changes in mortality over calendar years, such as those driven by medical advancements, public health interventions, or historical events like wars and pandemics. The trend shows an

apparent long-term decline, consistent with improved living standards and healthcare in Sweden over the past few centuries.

In contrast, the cohort effect $\hat{\delta}^f_{t-x}$, shown in Figure 10, captures systematic deviations in mortality risk associated with year of birth, after adjusting for age and period. The curve suggests that individuals born in the late 1800s and early 1900s experienced relatively higher mortality risk across their lifespans, while more recent cohorts exhibit lower mortality.

Together with the age effects summarized in Table 6, these plots offer a complete decomposition of mortality trends under the APC framework.

Table 7: Mean Absolute Percentage Error (MAPE,%) for ages 20-90 when the model is trained on three historical windows (1751-2000, 1850-2000, and 1950-2000) and projected to 2001-2023.

	MAPE (%)		
AgeGroup	**1751-2000**	**1850-2000**	**1950-2000**
20-24	11.60	12.50	8.62
25-29	10.30	10.20	8.27
30-34	13.30	14.50	9.89
35-39	15.50	15.70	11.40
40-44	14.80	14.60	10.30
45-49	19.10	19.00	13.00
50-54	17.70	15.50	13.60
55-59	20.30	20.40	12.70
60-64	20.80	20.90	13.40
65-69	21.90	20.20	12.70
70-74	25.80	26.60	15.30
75-79	24.50	21.60	17.00
80-84	30.20	28.30	13.00
85-90	29.50	27.60	14.80

Table 7 summarizes the out-of-sample forecast accuracy for an Age-Period-Cohort (APC) model trained on three different historical windows (1751-2000, 1850-2000, and 1950-2000). We measure performance using the mean absolute percentage error (MAPE) for mortality forecasts over the 2001-2023 horizon. The table is arranged by 5-year age groups (from 20-24 through 85-90), allowing us to see how well each training period predicts different segments of the adult lifespan.

The 1950-2000 model yields the lowest MAPE for most age groups, sometimes falling below 10%, especially in the 20-34 range. These values suggest that focusing on more recent data (the last 50 years

before 2000) can increase the predictive power of the APC model for younger to middle-aged cohorts, likely reflecting the dominant medical and social changes in the latter half of the 20th century. By contrast, while more historically comprehensive, the 1751-2000 and 1850-2000 windows tend to show higher errors for most ages. Including older mortality data brings in historical events and patterns (like pandemics or highly high mortality in specific periods) that may no longer accurately reflect modern conditions, leading to worse fits when forecasting past 2000.

Looking at specific age groups, the 45-59 and 70-79 brackets in the longer windows (1751-2000 or 1850-2000) typically experience MAPE values exceeding 15- 20%, while the MAPE of the 1950-2000 model remains closer to 10-15% for those same ages. Interestingly, for the oldest ages (80-90), the 1950-2000 model also achieves lower MAPE than the earlier training sets. This indicates that a more recent training window better captures advances in elderly care and life-extending medical interventions. Thus, although older data can help smooth random fluctuations, using an excessively long historical span may weaken more relevant trends that emerged in modern decades.

4.3.2 Forecasting 2024-2050

To generate these forecasts, we adopt a Poisson-regression formulation of the Age-Period-Cohort (APC) model [21]:

$$\log(D_{x,t}) \;=\; \log(E_{x,t}) \;+\; \alpha_x \;+\; \gamma_t \;+\; \delta_{t-x},$$

since

$$\log(m_{x,t}) \;=\; \log\!\Big(\tfrac{D_{x,t}}{E_{x,t}}\Big) \;=\; \log(D_{x,t}) - \log(E_{x,t}).$$

In other words, one can treat $\log(E_{x,t})$ as an offset in a Poisson regression for the death counts $D_{x,t}$.

We trained the model using data from 1950 to 2023 for people aged 20 to 84, which provides historical patterns of how mortality evolves by age, calendar year, and generational cohort. To forecast the cohort effects δ_{t-x}, we assume they follow an ARIMA$(1,1,0)$ with drift:

$$\delta_{t-x} \;=\; \delta_{t-x-1} \;+\; \phi\,(\delta_{t-x-1} \;-\; \delta_{t-x-2}) \;+\; c + \;\varepsilon_{t-x},$$

which was earlier described in Section 2.4.3.

To forecast the period effects γ_t, we employ a similar ARIMA$(1,1,0)$ structure. In the `StMoMo` package output, this is labeled as *mrwd*

(mean-reverting random walk with drift), but more precisely, it means that

$$\Delta\gamma_t \;=\; \phi\,\Delta\gamma_{t-1} + c \;+\; \varepsilon_t,$$

so that the changes $\Delta\gamma_t$ gradually revert toward zero if $|\phi| < 1$. We let $\Delta\gamma_t = \gamma_t - \gamma_{t-1}$ denote the first difference of the period effect. While this lowers the constant drift, the level γ_t is still non-stationary rather than reverting to a fixed mean. We obtain the forecasted values of γ_t from the last observed year by projecting $\Delta\gamma_t$ forward.

Finally, to generate the complete mortality forecasts, we combine the fitted age effects $\hat{\alpha}_x$ with the projected period and cohort components:

$$m_{x,t} \;=\; \exp\!\Big(\hat{\alpha}_x + \hat{\gamma}_t + \hat{\delta}_{t-x}\Big) \quad \text{for } t = 2024, \ldots, 2050.$$

This way, the APC model's decomposition of age, period, and cohort effects can be used to estimate future mortality rates under modern data-driven time-series assumptions.

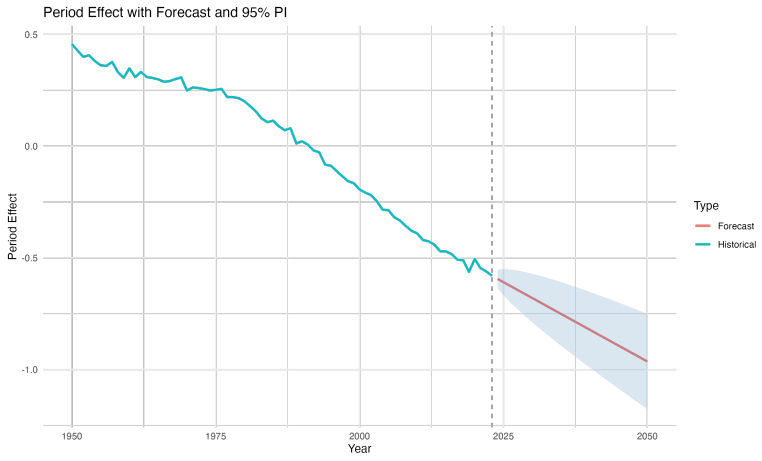

Figure 11: Forecasted period effects (γ_t) for 2024-2050, estimated using data from 1950-2023 with 95% Prediction Interval.

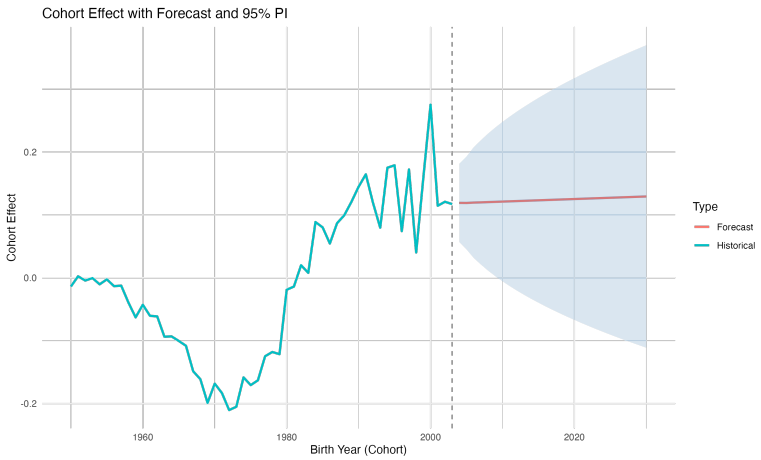

Figure 12: Forecasted cohort effects (δ_{t-x}) for 2024-2050, estimated using data from 1950-2023 with 95% PI.

Now that we have forecasted period and cohort effects, we use these components to calculate the projected mortality rates. As usual, we present the results in five-year age spans, where we average the single-year estimates within each bin. This aggregation provides a clearer view and improves interpretability [22].

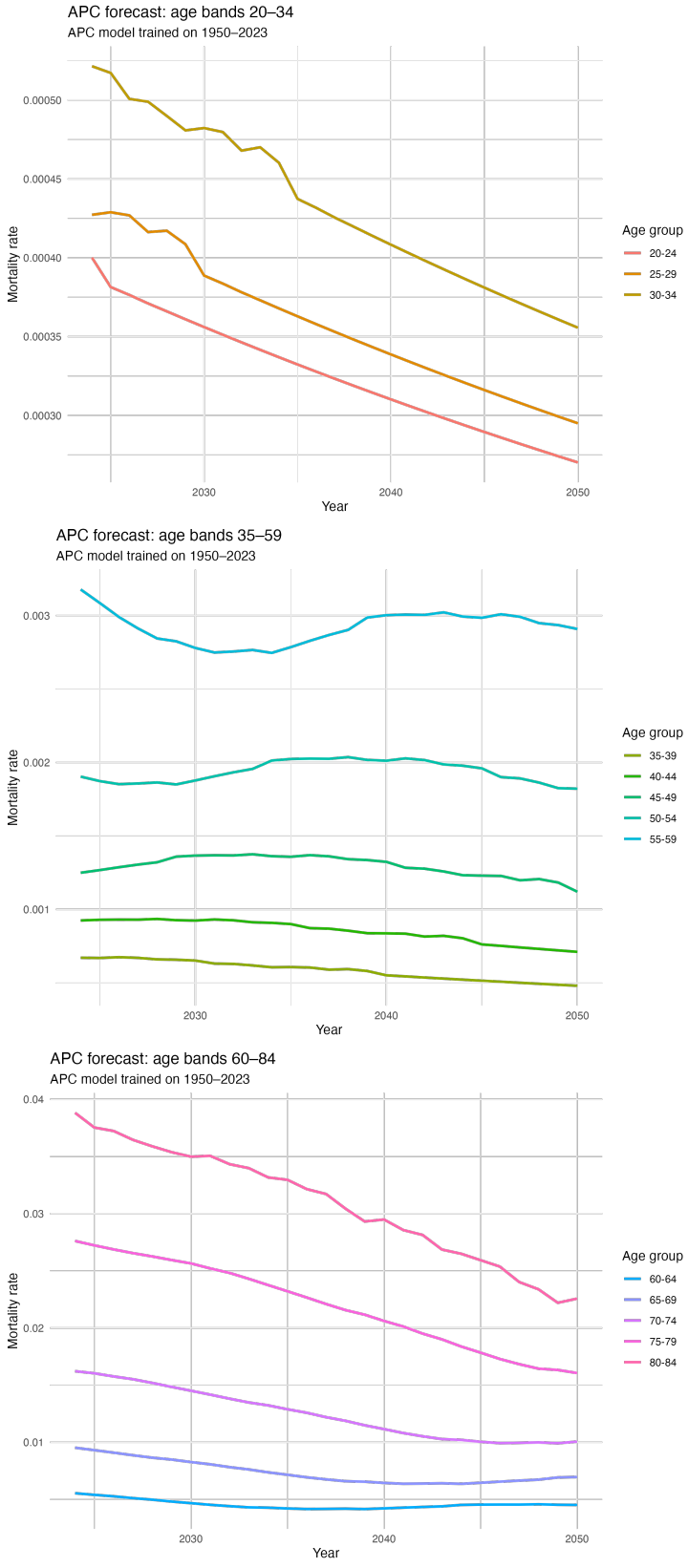

Figure 13: Forecasted mortality rates (2024-2050) from an Age-Period-Cohort (APC) model trained on 1950-2023 data, grouped in five-year age spans.

Figure 13 shows a consistent downward trend in mortality rates across all age groups, reflecting the overall progress in longevity. As age increases (e.g. 70-74, 75-79, 80-84), the baseline mortality rate is higher but gradually improves. Lower ages (20-24, 25-29, 30-34) exhibit very low, almost identical rates. These forecasts reflect the APC model's assumption that recent period effects and cohort trends continue meaning that older generations benefit from ongoing medical and social improvements, and that new cohorts retain their historically lower baseline mortality through the forecast horizon.

Similar to the Lee–Carter forecast, the youngest age groups (20–34) decline fastest. Low death counts at those ages make the fitted rates more volatile, which can exaggerate the downward slope. In addition, Figure 11 shows that the period effect γ_t has a strong negative drift, pulling every age downward year after year. Figure 12 shows that the cohorts aged 20–34 in 2024 (birth years > 1990) have the highest estimated δ_c values, giving them an especially low starting baseline and driving the continued decline in mortality.

In summary, the APC model leverages the separate influences of age, period, and cohort to produce age-specific mortality forecasts. Trained on data from 1950 to 2023, it projects continued reductions in adult mortality risk from 2024 to 2050, particularly among older cohorts, consistent with the improvements observed throughout the latter half of the 20^{th} century and into the early 21^{st}.

4.4 Comparison of Lee Carter vs APC

The Lee-Carter and Age-Period-Cohort (APC) models effectively capture long-term mortality trends but do so in fundamentally different ways. The Lee-Carter model assumes that a single time trend k_t drives mortality changes across all ages. In contrast, the APC model decomposes mortality into age, period, and cohort effects, allowing for a more flexible interpretation of generational patterns.

Out-of-sample evaluations (2001-2023) show that the Lee-Carter model performs best in the mid-adult range (ages 30-50), where mortality patterns are relatively stable. For example, when trained on 1751-2000 data, the MAPE for ages 60-64 is as low as 9.13%. However, forecast errors increase significantly for older ages (e.g. 70+), where the model tends to overestimate or underestimate late-life mortality.

In contrast, when trained on recent data (1950-2000), the APC model achieves lower MAPE scores, particularly for young and mid-adult age groups. For instance, the MAPE drops to 8.27% for the 25-

29 age group when focusing on recent trends, outperforming the model trained on a longer historical window. When trained on contemporary data, the APC model yields more reliable predictions for the oldest age groups (80-90), likely due to structural changes in elderly mortality driven by medical progress.

A key difference between the two models lies in how their time components are extrapolated. In our implementation, the Lee-Carter time trend k_t is modeled using a **random walk with drift**, corresponding to an **ARIMA(0,1,0)** model. This choice reflects the original assumption in the Lee-Carter framework that mortality improvements evolve gradually and persistently over time without reverting to a fixed level.

In contrast, the APC model forecasts the period and cohort components (γ_t and δ_{t-x}) using separate **ARIMA models**. In particular, we model the cohort effect using an **ARIMA(1,1,0) with drift**, which introduces mean-reverting behavior via the autoregressive term. This structure allows the model to adjust flexibly to past cohort-specific mortality fluctuations. However, it also increases the model's sensitivity to historical variation, leading to wider prediction intervals in long-term forecasts. The use of ARIMA(1,1,0) is motivated by cohort effects often fluctuate more erratically and exhibit generational reversals, which a simple random walk may fail to capture.

The APC model's explicit inclusion of cohort effects makes it better suited to capture long-term generational differences in mortality risk. However, the error in forecasting cohort effects is substantial, as reflected by the wide prediction intervals in later years, seen in figure 12. This uncertainty becomes especially pronounced toward the end of the forecast horizon (2040-2050), where projections are more sensitive to variation in past cohort trends.

The Lee-Carter model is simpler and more interpretable, with a straightforward decomposition into age-specific levels and a shared time trend. Its forecasts are smoother and generally more stable, particularly when a long historical series is used for training. However, it may fail to capture nonlinear or cohort-specific changes that the APC model can incorporate more effectively.

Lee and Carter [6] stated that their age-period model works the best when trained on 30-90 years, so most studies restrict the training window to a few decades. In our Swedish data, the lowest out-of-sample

error for the Lee-Carter model occurs when we train on the entire set. Similarly, for the Age-Period-Cohort (APC) model one might expect the opposite, that more data is needed due to its complexity. From [23] a researcher survey was made in which they said that APC needs ideally at least 50 years, or long enough for multiple generations to advance through the same life stages. It might seem a bit counterintuitive to go against the recommendations, but this is where we found the most success.

In summary, while both models are valuable tools for mortality forecasting, choosing between them involves a tradeoff between simplicity and flexibility. The Lee-Carter model is robust and interpretable but may miss cohort-driven dynamics. In contrast, the APC model captures more complex structures at the cost of increased uncertainty in long-term projections.

4.4.1 Comparison of different forecasting periods

Another way to compare the Lee-Carter and Age-Period-Cohort (APC) models is to evaluate them over different forecast periods. The motivation is that, in theory, Lee-Carter tends to perform better for shorter forecast periods because it relies on a single time-varying factor (k_t), which often suffices to model stable, gradual changes in mortality. In contrast, the APC model is more flexible and can capture nonlinear patterns across age, period, and cohort dimensions, features that have become increasingly important over more extended forecast periods.

To illustrate this difference, we consider two extreme forecasting periods: 5 and 50 years. In the 5-year forecasting span, we train the Lee-Carter and APC models from 1751-2018 and forecast 2019-2023. In the 50-year forecasting window, we train our models from 1751-1973 and forecast 1974-2023. We use the mean absolute percentage error (MAPE) as our primary metric for both APC and Lee-Carter. In addition, we report the mean absolute error (MAE) in Section 6, specifically in Tables 12 and 13, to offer a more detailed view of forecast accuracy. The procedure works exactly as in Sections 4.2.3 and 4.3.2.

Table 8: APC MAPE by Age Group for 50-Years vs. 5-Years Forecast

Age Group	**MAPE (50-Years)**	**MAPE (5-Years)**
20-24	15.9%	16.9%
25-29	18.3%	8.21%
30-34	22.7%	8.61%
35-39	23.1%	13.1%
40-44	28.5%	11.8%
45-49	32.8%	9.48%
50-54	37.1%	13.5%
55-59	39.7%	10.1%
60-64	41.9%	8.62%
65-69	42.9%	15.4%
70-74	42.9%	8.05%
75-79	41.7%	10.3%
80-84	39.5%	17.3%
85-90	43.3%	6.85%

Table 9: Lee-Carter Forecast MAPE by Age Group (50-Years vs. 5-Years)

Age Group	**MAPE (50-Years)**	**MAPE (5-Years)**
20-24	23.5%	32.9%
25-29	21.7%	7.19%
30-34	27.2%	12.7%
35-39	27.0%	31.8%
40-44	28.2%	7.04%
45-49	41.7%	9.72%
50-54	46.9%	31.3%
55-59	56.6%	5.80%
60-64	57.4%	19.3%
65-69	69.0%	23.0%
70-74	70.2%	5.95%
75-79	77.0%	25.1%
80-84	76.5%	18.1%
85-90	85.5%	9.52%

Table 8 shows the APC model's MAPE for 50-year versus 5-year forecasts. As anticipated, the extended forecast (50 years) yields higher errors (e.g. surpassing 40% MAPE for ages above 60), but it remains within a 15-43% range across all age groups. Meanwhile, the shorter (5 year) MAPE is substantially lower for most ages, consistent with fewer structural changes appearing in a brief forecast window.

Table 9 presents the corresponding Lee-Carter results. In the longer

horizon, Lee-Carter often produces higher MAPE than APC (e.g. reaching 85.5% in the 85-90 group), suggesting that a single k_t parameter may struggle to capture multi-decade shifts. Conversely, for the shorter horizon, Lee-Carter does show relatively low MAPE in select age groups (e.g. 5.80% for 55-59) and indeed outperforms APC in some instances.

According to [24], a MAPE < 10 % is "highly accurate", 10–20 % "good", 20–50 % "reasonable", and > 50 % "inaccurate". In our 50-year forecast, several Lee–Carter MAPE values would be judged inaccurate, whereas the APC model remains within the "reasonable" group throughout. However, [24] does not specify what the forecasting horizon should be. This is very important to understand, since the longer you forecast, the harder it is to forecast accurately.

4.4.2 Graphical out-of-sample comparison

In Sections 4.2.2 and 4.3.1 we compared three different training windows and found that the optimal periods are 1751–2000 for Lee-Carter and 1950–2000 for APC. To complement the mean absolute percentage error (MAPE) tables with a visual perspective, Figure 14 contrasts the forecasts from both models with the observed mortality rates for ages $20, 40, 60$ and 80 during 2001-2023.

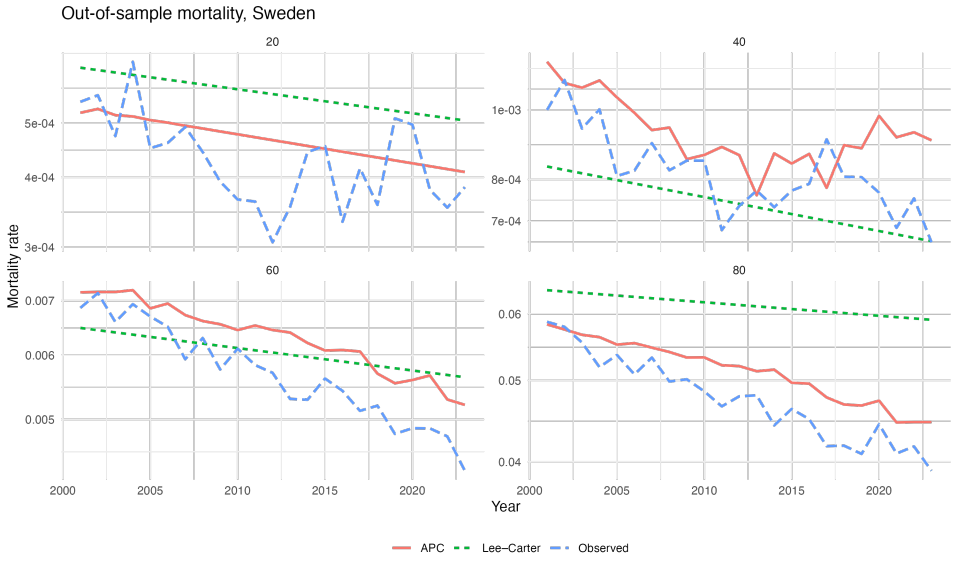

Figure 14: Out-of-sample mortality forecasts: Lee–Carter vs APC for ages $20, 40, 60$ and 80, 2001–2023

Figure 14 is consistent with the MAPE values in Tables 5 and 7. We see that Lee-Carter works best for the mid-aged group (40 and 60), which can be explained by the linear trend that k_t manages to capture. However, when it comes to younger and older ages, it cannot handle shifts the way APC does. Due to the complexity of APC and the inclusion of the cohort effect, the APC can handle shifts in mortality that Lee-Carter might miss.

5 Discussion

The Lee-Carter model, built around a single time-varying parameter k_t, offers simplicity, interpretability, and stable performance when trained on long historical data. Our results confirm that training on an entire historical window (1751–2000) yields the lowest out-of-sample errors (MAPE) overall, especially for mid-aged groups (e.g. ages 40–69), with an exceptionally low error at ages 60–64 (9.13% MAPE).

The APC model decomposes mortality into separate age, period, and cohort effects, capturing more complex generational patterns. This flexibility comes at the cost of more significant forecast error, especially evident in the widening prediction interval of the cohort component (Figure 13). Compared to the Lee-Carter method, the APC model achieves better predictive performance when trained on more recent data (1950–2000), particularly for younger and older age groups. For instance, the 25–29 age group achieves a MAPE of 8.27%, and even the 80–84 age group sees a dramatic improvement in forecast accuracy relative to training on the historical dataset. The success of these smaller training sets suggests that the APC model benefits from capturing modern medical and societal trends that disproportionately affect specific cohorts.

An important methodological distinction is the choice of ARIMA models for forecasting the time components. In the Lee-Carter model, k_t follows a random walk with drift (equivalent to ARIMA(0,1,0)), which assumes that mortality improvements evolve smoothly over time. Conversely, the APC model uses ARIMA(1,1,0) processes for the period and cohort effects, allowing for slight mean reversion and smoother generational transitions. This choice is motivated by the unpredictable, nonlinear shifts in cohort-specific mortality that a pure random walk might not capture. While this makes the APC model more flexible in capturing historical volatility, it also increases its sensitivity to past fluctuations.

Another key finding is that the optimal training period differs between the two models. The Lee-Carter model performs best when leveraging a long data span, taking advantage of the stability of its single-time index across centuries. On the other hand, the APC model tends to benefit from being trained on more recent data that better reflect ongoing structural changes in mortality, especially among elderly cohorts where rapid healthcare improvements have had profound effects.

In Section 4.4.1 we compared the two models over extreme horizons: a short-term 5-year forecast and a long-term 50-year forecast. Lee-Carter records its smallest errors at the 5-year horizon, even outperforming APC in a handful of middle-age groups (25-29, 40-44, 55-59, and 70-74). Nevertheless, APC has the lower mean error across all ages at both horizons and is markedly more accurate over 50 years. This behaviour is consistent with Lee-Carter's single-factor structure, which captures smooth near-term period trends, whereas APC's additional cohort term pays off as structural changes accumulate over decades.

Since we concluded that neither model outright dominates the other, each has strengths in different age ranges and data parts. APC is more robust overall, while Lee-Carter can more accurately capture recent trends at certain ages. This means that forecasters or actuaries might choose APC for very long-term projections (favoring its complexity), but use Lee-Carter for short-to medium-term forecasts for specific age groups.

5.1 Possible future enhancements

Working on this thesis was rewarding because of its complexity. There were always several paths that could be taken to explore, improve, and extend the analysis of this thesis. However, this also results in a lot of important observations being left out. We will now discuss the most important ones that got left out and would be interesting to work on further.

Throughout this thesis, the mortality rates were grouped into five-year age groups (e.g. 20-24, 25-29, ...) rather than modeled at individual age levels. The reasoning behind this is to improve the clarity of visualizations and reduce the number of plots and tables. Although this might be nice for the reader, it introduces several trade-offs that affect the precision of the analysis.

By grouping ages, we effectively compute the average mortality within each interval, which can hide important differences between individual ages. For instance, the mortality pattern at age 20 may differ substantially from that at age 24, particularly in younger or older populations where risks can change quickly from year to year. As a result, some of the finer age-related dynamics might go lost in this process.

Even though grouped data simplifies interpretation and aligns with many other demographic reporting standards, future work could explore whether narrower age groups or even single-year modeling might reveal more detailed results, especially in age intervals where mortality

changes rapidly. On the other hand, finer modeling increases the dimensionality and may require regularization or smoothing techniques to avoid overfitting, especially in sparse age ranges.

Thus, while five-year grouping provides a reasonable balance between interpretability and complexity, it is important to acknowledge its limitations and consider whether specific age intervals might benefit from further subdivisions depending on the application.

Another important consideration is the choice of training window. As shown in Section 4.2.2, Lee-Carter benefits from longer historical sequences, whereas APC performs best when trained on more recent data. Varying the training periods and evaluating performance could help determine more optimal horizons for different models and age groups.

Expanding the dataset could also enhance model performance. Given the demographic similarities between these populations, incorporating mortality data from neighboring Scandinavian countries such as Norway, Denmark, or Finland may provide valuable additional structure. By including our Scandinavian neighbors, we could use multi-population extensions of the Lee-Carter and APC frameworks, leading to more precise and realistic forecasts.

As discussed in Section 4.2.5, the model performance would improve significantly by removing outliers. While this improves forecast stability and fit, it is important to recognize that such outliers are a part of history. Incidents like pandemics or wars, though rare, show meaningful impacts on mortality trends and should not be dismissed lightly. A recent example is the COVID-19 pandemic, a significant outlier in future mortality analyses. This thesis aims to provide a thoughtful and historically grounded view of mortality development. Therefore, we have included all available years and central events dating back to 1751, when mortality statistics in Sweden first started to be recorded for real. Including all years allows the analysis to reflect the general long-term trends and the extreme events shaping them.

A natural extension of the parametric bootstrap in Section 4.2.4 is to apply the same procedure to the Age–Period–Cohort (APC) model and to look at the raw variation of $\frac{D_{x,t}}{E_{x,t}}$. It would be interesting to see how well the APC model handles scaled exposures. We expect it to "break" faster than Lee–Carter due to its added complexity and larger parameter set (period + cohort effects).

When comparing the variability of raw observed rates to the variability of modeled rates, we expect that raw rates appear more unstable, especially when $E_{x,t}$ is small. This is because raw rates reflect all the year-to-year random fluctuations in deaths, whereas a model like Lee–Carter or APC are fitted to impose a smoother structure.

Due to time and computational constraints, this analysis was not implemented. However, it would provide valuable insights into the robustness of APC's period and cohort effects and the variation in our mortality rates.

6 Appendix

6.1 Female Lee Carter analysis

We evaluated the prediction ability of the Poisson Lee-Carter model using three training windows: *1751-2000*, *1850-2000*, and *1900-2000*. After fitting the model in each case, we projected mortality for ages 20-80 from 2001 through 2023. Table 10 presents the root mean square error (RMSE) between predicted and observed mortality rates in these out-of-sample years, with lower values indicating closer alignment with reality.

Table 10: RMSE of Poisson Lee-Carter forecasts (2001-2023) by training window and age.

Age	**1751-2000**	**1850-2000**	**1900-2000**
20	0.000108	0.000123	0.000200
30	0.0000773	0.000149	0.000219
40	0.000112	0.000213	0.000229
50	0.000342	0.000325	0.000291
60	0.000592	0.000505	0.000407
70	0.00159	0.00140	0.00120
80	0.0131	0.0106	0.00846

RMSE Overview.

- **1751-2000:** Ages **30** and **40** yield particularly low RMSE (e.g. 0.0000773 for age 30), while **80** reaches 0.0131, making it the worst within this window.
- **1850-2000:** Age **20** has the smallest RMSE (0.000123), while age **80** is again the largest at 0.0106.
- **1900-2000:** Age **20** (0.000200) outperforms the others, while age **80** (0.00846) is the most difficult to predict.

6.2 Forecasting between 30-90 years according to Lee-Carter

Table 11: Mean Absolute Percentage Error (MAPE) over 1981-2023 when training between 1950-1980.

Age Group	**MAPE (1950-1980)**
20-24	18.1
25-29	18.1
30-34	29.0
35-39	48.5
40-44	68.2
45-49	49.0
50-54	41.1
55-59	27.8
60-64	22.6
65-69	22.4
70-74	21.3
75-79	19.2
80-84	6.52
85-90	5.62

Lee and Carter recommend using a training window of 30-90 years for mortality forecasts, which prompted us to test a 30-year span. However, as shown in Table 11 (Appendix), the resulting MAPE values are worse than those in our main analysis. One likely reason is that a strictly 30-year window may not include enough historical variation to stabilize the estimated trends. Furthermore, a comparatively short training period can overemphasize recent anomalies or improvements (e.g.
breakthroughs in medical technology) and fail to capture broader long-term mortality patterns.

6.3 Lee-Carter vs. APC: MAE Comparison

Table 12: APC MAE by Age Group for 50-Year vs. 5-Year Forecast

Age Group	**MAE (50-Year)**	**MAE (5-Year)**
20-24	0.00261	0.0000806
25-29	0.00327	0.000294
30-34	0.00249	0.00245
35-39	0.00335	0.0000818
40-44	0.00392	0.000518
45-49	0.00613	0.00570
50-54	0.00594	0.000157
55-59	0.00894	0.000734
60-64	0.00816	0.00264
65-69	0.0108	0.000208
70-74	0.00905	0.000812
75-79	0.0120	0.00359
80-84	0.0100	0.000358
85-90	0.0130	0.00342

Table 13: Lee-Carter Forecast MAE by Age Group (50-Year vs. 5-Year)

Age Group	**MAE (50y)**	**MAE (5y)**
20-24	0.00281	0.000200
25-29	0.00482	0.000257
30-34	0.00459	0.00514
35-39	0.00699	0.000200
40-44	0.00589	0.000329
45-49	0.00905	0.00166
50-54	0.00790	0.000248
55-59	0.0111	0.000746
60-64	0.00924	0.00423
65-69	0.0132	0.000199
70-74	0.0113	0.00147
75-79	0.0149	0.00293
80-84	0.0124	0.000196
85-90	0.0167	0.00455

Tables 12 and 13 display the mean absolute error (MAE) for the Age-Period-Cohort (APC) and Lee-Carter (LC) models, respectively, when forecasting over a 5-year horizon vs. a 50-year horizon. These tables extend Section 4.4.1 discussion by showing the numerical details behind the 5-year and 50-year forecasts.

In both models, the 5-year horizon generally exhibits lower absolute errors than the 50-year horizon across most age groups, consistent with the idea that short-term forecasts are more reliable. However, there are specific age groups (significantly younger or older) where the models' MAE patterns differ substantially, highlighting their distinct sensitivities to different age segments.

For instance, the Lee-Carter model may yield low MAE for mid-adult ages (e.g. 20-44) when projecting 5 years ahead. In contrast, the APC model can excel for particular older-age cohorts if structural changes are captured more accurately. These differences underscore the broader conclusion that the choice of model and forecast horizon depends on which age ranges and time frames are most relevant to the analyst's goals.

7 References

References

[1] Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data, Second Edition*. Ch. 1, pages 5-8.

[2] Aalen, O., Borgan, O. and Gjessing, H. (2008). *Survival And Event History Analysis*. Ch. 3.1.1, page 71-72:

[3] Box, G. E. P., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. 5th ed. Ch. 4 and 5.

[4] Hyndman, J. R. and Athanasopoulos, G (2024). *Forecasting: Principles and Practice*. 3rd ed. Ch. 9.3 and 2nd ed. Ch. 8.7.

[5] Nau, R. (2020). *Introduction to ARIMA: nonseasonal models*

[6] Lee, R. D. and Carter, L. R. (1992). *Modeling and Forecasting U.S. Mortality*. Journal of the American Statistical Association, 87(419): 659-671.

[7] Brouhns, N., Denuit, M. and Vermunt, J (2002). *A Poisson Log-Bilinear Regression Approach to the Construction of Projected Lifetables*. Insurance: Mathematics and Economics, 31(3), 373–393.

[8] Eriksson, C (2020). *MULTI-POPULATION MORTALITY MODELS IN THE LEE-CARTER FRAMEWORK - AN EMPIRICAL EVALUATION ON SWEDEN'S 21 COUNTIES*. Page 7.

[9] Fosse, E. and Winship, C (2019). *Analysing Age-Period-Cohort Data: A Review and Critique*. pages 3-13.

[10] Villegas, A. M. Millossovich, P. and Kaishev, V. K. *StMoMo: An R Package for Stochastic Mortality Modeling*. Ch. 3.1 and 3.3.

[11] Kuang, D. Nielsen, B. and Nielsen, J.P (2008). *Forecasting with the age-period-cohort model and the extended chain-ladder model*. Biometrika, 95(4), 987–991.

[12] Ohlsson, E. and Johansson, B. (2010). *Non-Life Insurance Pricing with Generalized Linear Models*. Ch. 3, page 40

[13] Hyndman, J. R. and Athanasopoulos, G. (2024). *Forecasting: Principles and Practice*. 3rd ed. Ch. 5.8.

[14] Sundberg, R. (2021). *Lineära statistiska modeller*. Ch 6.8, page 275.

[15] Blomberg, N (2022). *A Comparison Between Different Stepwise Regression Models To Predict Football Games*. Page 10-11.

[16] *https://www.mortality.org/Home/Index* Human Mortality Database

[17] Currie, I. D (2016). *On fitting generalized linear and non-linear models of mortality*. Scandinavian Actuarial Journal, 356-383.

[18] Vries. R. D. (2011). *SO-rummet. Spanska Sjukan*. `https://www.so-rummet.se/kategorier/spanska-sjukan#`

[19] R. D. Lee and L. R. Carter. (1992). *Modeling and Forecasting U.S. Mortality. Journal of the American Statistical Association*, vol. 87, no. 419, 1992, pp. 659–671.

[20] (2023). *Hjärt-Lungfonden. 1958 – Rätt rytm med pacemaker*. `https://www.hjart-lungfonden.se/forskning/stora-genombrott/pacemakern/`

[21] Holford, T. R. (1983). *The Estimation of Age, Period and Cohort Effects for Vital Rates*. Biometrics, 39(2), 311–324.

[22] Dolores, M. et.al (2014). *Inference and Forecasting in the Age–Period–Cohort Model with Unknown Exposure with an Application to Mesothelioma Mortality*. Journal of the Royal Statistical Society Series A: Statistics in Society, Volume 178, Issue 1, January 2015, Pages 29–55.

[23] Lau, A. and Kennedy, C (2023). *Assessing the effects of generation using age-period-cohort analysis*. Pew Research Center.

[24] Moreno, M.J.J. et.al (2013). *Using the R-MAPE index as a resistant measure of forecast accuracy.* Vol. 25, No. 4, Page 501

[25] Wang, J. Z (2007). *Fitting and Forecasting Mortality for Sweden: Applying the Lee-Carter Model*. pages 10-11.

[26] Willekens, F.J. and Baydar, N. (1984). *Age-period-cohort models for forecasting fertility*. Ch. 2, page 6-9.