# Modeling and forecasting mortality with economic, environmental and lifestyle variables

Matteo Dimai, Università degli Studi di Trieste, matteo.dimai@phd.units.it, Via Valerio n. 4/1 - 34127 Trieste (TS), Italy

05/09/2023

Traditional stochastic mortality models tend to extrapolate trends without exploring underlying drivers. Those that do link mortality with other variables usually limit themselves to GDP. This article introduces a novel stochastic mortality model that incorporates a broad spectrum of variables related to economic, environmental, and lifestyle factors to predict mortality. The model uses principal components derived from these variables in an extension of the Niu and Melenberg (2014) model and is applied to 37 countries from the Human Mortality Database. Model fit is superior to the Lee-Carter model for 18 countries and closely matches it in others. The forecasting accuracy of the proposed model improves on the Niu-Melenberg model for half of the countries studied under various jump-off years. Sustainable populations require an intricate understanding of the interplay between mortality and its determinants. The model is designed to facilitate scenario building and policy planning, including strategies for population sustainability. By examining a comprehensive array of variables, this model contributes to a holistic comprehension of population dynamics.

Keywords: mortality, GDP, environment, forecasting, tobacco, scenario planning

## Statements and Declarations

**Ethics approval and consent to participate**: Not applicable.

**Consent for publication**: Not applicable.

**Availability of data and materials**: The data that support the findings of this study are freely available online and from the author upon request. The R code used is available from the author upon request.

**Competing interests**: The author has no competing interests to declare that are relevant to the content of this article.

**Funding**: No funds, grants, or other support was received.

**Authors' contibutions**: Not applicable (sole author).

**Acknowledgements**: The author would like to thank prof. Pietro Millossovich and prof. Susanna Zaccarin for their helpful comments and the anonymous reviewers for their help in improving the article.

# 1 Introduction

Over the past century, the average human lifespan has increased significantly. In 1913, the global life expectancy was estimated to be 34.1 years, while in 2001 it increased to 66.6 years, and by 2015, it had further increased to 71.8 years (Riley, 2005; Wang et al., 2016). Forecasting the evolution of mortality rates is very important for actuarial practice, as well as for health care and pension systems in general, and so forecasts should be carefully assessed.

Yet, forecasts are based on past and present trends and a fundamental uncertainty is the extent to which these trends will remain valid in the future. Climate change and the economic effects of the green transition create an interplay of forces that cannot simply be ignored when forecasting mortality. Hence it is important to model uncertainty not only to the extent that it is expressed as variability around current trends, but also as to incorporate sources of variation that can affect, perhaps even reverse current trends. What drives trends in mortality and what would happen if those drivers changed over time? Human health depends on a large number of factors, including biological, environmental and social ones. These drivers may have reinforcing or opposing effects and are subject to varying degrees of uncertainty. This paper aims to develop a model that can incorporate a wide range of covariates and be used to forecast mortality under different scenarios. The main assumption is that mortality and the covariates do not diverge in the long run, without requiring causality assumptions. By broadening the scope of covariates used in stochastic mortality modeling and developing an interpretable model, this study aims to provide more accurate forecasts of mortality rates.

Previous research in the field of stochastic mortality models incorporating external variables is relatively scarce and usually only includes GDP as the driver of mortality reductions. Niu and Melenberg (2014) and Seklecka et al. (2019) approach the issue with a single-population model, while Boonen and Li (2017) presents a multi-population model. Dutton et al. (2020) extends Seklecka et al. (2019) by including the effect of temperature anomalies on mortality. On the other hand, there is a vast literature that forecasts mortality using a multiplicity of variables, including the ones in this study, which all have a well-known link with mortality. However, to the author's knowledge this it is the first study in which they are included together in a stochastic mortality model.

The contribution of this study is to show how broadening the scope of external variables included in stochastic mortality models can improve forecasts compared to Niu and Melenberg (2014) and at the same time allow actuarial practitioners and policymakers to use interpretable models to assess the impact of different scenarios on mortality. Additionally, the study will explore the extent to which different variables have a disparate impact on mortality in different countries, highlighting the need to tailor the analysis to each country's specific characteristics and history.

The rest of the paper is organized as follows. Section 2 presents a review of the literature on stochastic mortality models, with a focus on the ways to include external variables in the modeling process. The single-population model with external variables (SEV) is introduced in section 3 along with details about the fitting process, about the variables and an analysis of the characteristics of the time series used. The results are presented in section 4, with a focus on 5 countries and a general overview of goodness of fit and forecasting performance for all countries considered. Section 5 concludes with a discussion of the findings.

## 2 Literature review

The approaches to forecast mortality are generally classified into three categories, namely the extrapolation, explanation and the expectation ones (Stoeldraijer et al., 2013). Extrapolation approaches assume that existing trends and patterns in mortality rates by age are regular enough to continue into the future. Explanation approaches model future mortality with exogenous variables that have a known link to mortality, like smoking and lung cancer, using structural or epidemiological equations. The expectation approach instead incorporates expert opinions regarding various aspects of mortality. The three approaches can coexist in a single model and the boundaries between the approaches can be blurred. Using external covariates to model mortality is close to the explanation approach. However, when applied without explicitly modeling the dependency structure between variables, this results in a model closer to extrapolative ones. The present work hence draws partly from the explanation and partly from the extrapolation approach, which is the one stochastic modeling has mostly concentrated on (Cairns et al., 2011). A common approach to extrapolative modeling is to extend past trends into the future by fitting linear trends for log-mortality rates with a temporal component that can be specified, for example, through a random walk with a linear or quadratic drift. These models may be accurate in the short run and backtesting, but they are unable to identify turning points and are not useful for assessing mortality under various development and policy scenarios. Even when models allow for the incorporation of expert judgment to set mortality targets (e.g., Boumezoued et al. (2019)), these targets are essentially set exogenously.

The link between economic development and mortality has been widely studied over the past decades. Preston (1975) described the association between life expectancy and per capita income almost 50 years ago. Further studies established both positive (Brenner (2005), Birchenall (2007)) and negative (Tapia Granados (2008), Tapia Granados and Ionides (2011), Tapia Granados and Ionides (2017)) effects of economic growth.

There are examples of stochastic mortality models that incorporate economic variables in the context of extrapolative models based on Lee and Carter (1992). Hanewald (2011) describes a relationship between the latent factor of the Lee-Carter model, GDP and unemployment. Building on that, Niu and Melenberg (2014) propose a more general model that allows for multiple latent and multiple exogenous factors and offers an application using GDP as a predictive variable. An extension to the multipopulation case is provided by Boonen and Li (2017), with a model that allows for additional exogenous variables and where the latent factor of the Lee-Carter model is dropped entirely. Most notably, they focus on multiple groups of populations, including post-Soviet countries who experienced a mortality increase after the drop in GDP due to the dissolution of the Soviet Union: the temporary increase in mortality in those countries before GDP recovered solidifies the relationship between GDP and mortality. Bozzo et al. (2021) represents a further development applied to the mortality between regions in Italy.

Other models lend themselves to the inclusion of economic variables. The model from O'Hare and Li (2012), which extends the model of Plat (2009) and allows for additional terms to estimate young mortality, is modified by Seklecka et al. (2019) by using the correlation between GDP and mortality. Dutton et al. (2020) took this further by multiplying a term of the model by the correlation between temperature anomalies and mortality, emphasizing the importance of environmental effects. Non-economic explanatory variables have also been included in some models. For example, French and O'Hare (2014), building on King and Soneji (2011) and on the literature that links mortality to lifestyle and dietary variables, present a model with GDP, health care expenditures, tobacco and alcohol consumption, fat intake and fruit and vegetables consumption.

More recently, Li and Shi (2021) propose a global vector auto-regressive (GVAR) approach that can model mortality rates for a large number of populations with the inclusion of global factors. These could be external factors like GDP or other covariates. In their application, though, the authors only employ the average of the mortality rates of the 15 countries they include, as a proxy for the global advancement of medical treatments.

An important challenge when modeling mortality by including multiple economic, social, environmental and technological factors is the lack of micro-level data. Ideally, death rates would need to be available for every possible subpopulation defined by every combination of the factors one wishes to study, or even at the individual level in case of continuous factors, covering several decades and the whole population. Although there are some studies working with individual data (i.e. Cairns et al. (2019)), the time frame of available data and the selection of covariates is usually limited, resulting in models that apply only for a limited age range. Therefore, it is necessary to use factors that can be thought of having a wider, general effect on a population. Since the relationship between the external factor and mortality is rarely direct, much less known with certainty, these models lie somewhere between pure extrapolation models and explanation models.

An alternative approach to modeling mortality with external factors is to divide the population into subpopulations based on a specific covariate, such as an affluence or development index, and use a multipopulation model for the different subpopulations, ensuring coherent forecasts. This method does not explicitly include the covariate in the model, but instead forecasts mortality for each subpopulation separately, assuming that the subdivision of the population according to the covariate is stable and accurate. Like regions in a state (i.e. Bozzo et al. (2021) and Danesi et al. (2015) for two examples with Italian regions), other groupings of the population at the subnational level can be reasonably believed to have a converging mortality pattern in the long run, while allowing for divergence in the short run. For ranked groups, a desirable outcome is for groups to preserve their ordering, that is, to avoid crossovers in estimates of mortality rates.

An example of this approach is given by Villegas and Haberman (2014), where a composite deprivation index is used to rank small areas in England, classify them based on their rank and track their mortality counts. The authors show how less deprived quintiles had lower mortality throughout the period and also experienced faster mortality declines than more deprived quintiles, leading to a widening of mortality differentials by socioeconomic status.

# 3 Methodology and data

The availability and comparability of data across multiple countries is a main concern of this study, which meant that data sources have been carefully selected in order to ensure comparability and both the countries analyzed and the variables included reflect the availability of high-quality, comparable data.

The mortality data used in this article is limited to males and has been sourced from the Human Mortality Database (HMD), imported into R through the `HMDHFDplus` package. All 42 available countries were included. The choice of the external variables has been more problematic. On one hand, plausible covariates are variables that have already shown a link with mortality at the individual level or in forecasting or non-forecasting population-wide models. These include GDP, health care spending, other affluence measures, education, up to lifestyle variables like alcohol and tobacco consumption, obesity, marital/cohabitation status etc. On the other hand, comparability required that the external variables be chosen among those available from reliable, official data sources, like WHO or FAO databases, Penn World Tables and others, with time series extending as far back as the 1970s. The model estimates males and females separately and although only males are discussed for brevity, all variables, including mortality data, are available for females as well. Model estimation has been carried out with a version of the `StMoMo` R package modified by the author to allow for the inclusion of external variables.

## 3.1 Modeling methodology

The model proposed in this article is based on the Lee-Carter model, where the latent factor is substituted by one or more factors obtained from external variables through principal components analysis (PCA). The methodology is a mix of Niu and Melenberg (2014), Boonen and Li (2017) and French and O'Hare (2014), and can be summarized as follows:

1. Select possible covariates based on availability of data for a wide set of countries, an extended number of years and from reputable data sources. In addition to GDP and other affluence-related variables, consider variables related to environmental effects (i.e. air quality) and to lifestyle choices (i.e. alcohol consumption) that have a strong empirical link to mortality;
2. Perform tests on the possible covariates to assess their characteristics, i.e. determine whether they are stationary, non-stationary or stationary with structural breaks. Explore the long-term relationships between them and mortality, i.e. by checking whether cointegration relationships exist and are stable between countries;
3. For all countries with available data, fit the following single-population model, based on the model specified by Niu and Melenberg (2014):
$$log(m_{x,t}) = a_x + \sum_{j=1}^{J} b_{j,x}k_{j,t} + \sum_{l=1}^{L} c_{l,x}g_{l,t} + \epsilon_{x,t},$$
with $J$ age-period terms ($J = 0$ or $J = 1$ in the applications) and $L$ orthogonal external factors $g_{l,t}$. The external factors are combinations of multiple external variables, in order to reduce dimensionality and solve identification issues akin to the use of principal components in Boonen and Li (2017).

The proposed approach tries to minimize assumptions about the regularity of mortality rates and of the covariates. The main assumption instead is that there are stable long-term relationships between covariates and mortality, even if the series themselves aren't stationary or exhibit structural breaks. To this end, the cointegration analysis at step 2 investigates the existence of said long-term relationships: the cointegration coefficients themselves, though, aren't used in the model. The assumption of a long-term relationship between variables is a crucial one if the resulting model is to be used in scenario planning where forecasts are required under deviations from current trends.

## 3.2 Parameter estimation

The fitting procedure of the SEV models draws upon the Generalized Age-Period-Cohort (GAPC) models as implemented in the `StMoMo` R package (Villegas et al., 2018), which is itself based on the `gnm` R package, drawing from Niu and Melenberg (2014) as well. The `StMoMo` package has been extended in order to allow for external terms in the fitting procedure. Even though `StMoMo` allows offsetting terms to be included in the model, they are wholly external to the fitting process and therefore age loadings can't be fit for them without modifying the package. The following implementation bridges this gap in capability.

The SEV model for $J$ age-period terms and $L$ external terms is as follows:

$$log(m_{x,t}) = a_x + \sum_{j=1}^{J} b_{j,x}k_{j,t} + \sum_{l=1}^{L} c_{l,x}g_{l,t} + \epsilon_{x,t}.$$

The GAPC models model deaths instead of death rates. This is equivalent to fitting death rates and therefore death rates will be used in the following instead of deaths and exposures for compactness.

The terms $g_{l,t}$ are linear combinations of the $O$ external variables $h_{o,t}$, with $L < O$. In the applications, $J = 1$.

The fitting algorithm is as follows:

- Obtain yearly deaths by age $\mathrm{D}_{x,t}$, their correspondent exposures $\mathrm{E}_{x,t}$ (the ratio of these two quantities is equivalent to death rates $\mathrm{m}_{x,t}$) and $O$ external variables $h_{o,t}$ for the given country;
- Perform a singular value decomposition on the scaled matrix of $h_{o,t}$ external variables, retain the $L$ components which explain a share of variance larger than a set threshold, up to a given maximum of components, denote them with $g_{l,t}$;
- Estimate the model with `gnm`, obtaining the quantities $a_x$, $b_{j,x}$, $k_{j,t}$ and $c_{l,x}$.
- Fit the rates using the estimated parameters.
- Transform the parameters so that they satisfy the identifiability constraints.
- Fit the rates with the transformed parameters and check whether the transformation preserves the rates. If so, output the model.

The variance threshold, the maximum number of principal components and whether to include an age-period term are parameters set before fitting. The parameters, as is usual for mortality models based on the Lee-Carter model, are not identified without additional constraints. For example, setting $J = 1$, with $c^* \in \mathbb{R}^J$, $d^* \neq 0$ and $e^* \in \mathbb{R}^L$, it is possible to obtain, for the logarithm of the fitted rate $\mu_{x,t}$:

$$
\begin{aligned}
\mu_{x,t} &= a_x + b_x k_t + \sum_{l=1}^{L} c_{l,x} g_{l,t} \\
&= a_x + b_x k_t + \sum_{l=1}^{L} c_{l,x} g_{l,t} + \sum_{l=1}^{L} e_l b_x g_{l,t} - \sum_{l=1}^{L} e_l b_x g_{l,t} + b_x c^* - b_x c^* \\
&= (a_x - b_x c^*) + \frac{b_x}{d^*} d^* \left( k_t - \sum_{l=1}^{L} e_l^* g_{l,t} + c^* \right) + \sum_{l=1}^{L} \left( c_{l,x} + e_l^* b_x \right) g_{l,t} \\
&= \tilde{a}_x + \tilde{b}_x \tilde{k}_t + \sum_{l=1}^{L} \tilde{c}_{l,x} g_{l,t} \,,
\end{aligned}
$$

with

$$
\begin{aligned}
\tilde{a}_x &= a_x - b_x c^* \\
\tilde{b}_x &= \frac{b_x}{d^*} \\
\tilde{k}_t &= d^* \left( k_t - \sum_{l=1}^{L} e_l^* g_{l,t} + c^* \right) \\
\tilde{c}_{l,x} &= c_{l,x} + e_l^* b_x .
\end{aligned}
$$

It follows then that $c^*$, $d^*$ and $e_l^*$ need to be functions of the parameters such that, after the transformation, calculating $c^*$, $d^*$ and $e_l^*$ again yields $c^* = 0$, $d^* = 1$ and $e_l^* = 0$.

The following normalization constraints, based on Niu and Melenberg (2014) and Boonen and Li (2017) are proposed:

$$
\sum_{x=1}^{N} b_x = 1,
$$

$$
\sum_{t=1}^{T} k_t = 0,
$$

$$
k = (k_1, \ldots, k_T) \neq 0,
$$

$$
\sum_{t=1}^{T} k_t g_{l,t} = 0, \text{ for } l = 1, \ldots, L,
$$

with the last constraint describing that the sample covariance of $k_t$ and $g_{l,t}$ is 0. These correspond to the following transformations:

$$
\begin{aligned}
c^* &= -\frac{\sum_{t=1}^{T} k_t}{T} \\
d^* &= \sum_{x=1}^{N} b_x \\
e_l^* &= cov(g_{lt}, k_t)/var(g_{lt}) \text{ for } l = 1, \dots, L,
\end{aligned}
$$

with `cov()` and `var()` denoting sample covariance and variance, respectively. In order to demonstrate that the proposed constraints identify the model uniquely, the following theorem must be proven:

**Theorem 1 (Identification)**: Let $\mu = (\mu_{x,t})_{x=1,\dots,N,t=1,\dots,T} = \mu(\theta)$, where $\mu = \mu(\theta)$ satisfies $\mu_{x,t} = a_x + b_x k_t + \sum_{l=1}^{L} c_{l,x} g_{l,t}$ for some $\theta = \left((a_x)_{x=1,\dots,N}, (b_x)_{x=1,\dots,N}, (k_t)_{t=1,\dots,T}, (c_{l,x})_{x=1,\dots,N,l=1,\dots,L}\right)$. Then the parametrization $\theta^0$ satisfying the normalization constraints above satisfies the following:

- $\theta^0$ is a function of $\theta$.
- $\mu$ is a function of $\theta$ through $\theta^0$.
- The parametrization of $\mu$ by $\theta^0$ is exactly identified. That is, if $\theta^1 \neq \theta^2$ are two sets of parameters satisfying the normalization constraints, then $\mu(\theta^1) \neq \mu(\theta^2)$.

**Proof of Theorem 1**: The proof follows the ones in Niu and Melenberg (2014) and Boonen and Li (2017).

- For any $\theta$, use $\mu_{x,t} = (a_x - b_x c^*) + \frac{b_x}{d^*} d^* \left(k_t - \sum_{l=1}^{L} e_l^* g_{l,t} + c^*\right) + \sum_{l=1}^{L} \left(c_{l,x} + e_l^* b_x\right) g_{l,t}$ and construct $\theta^0$ by letting $d^* = \sum_{x=1}^{N} b_x$, $c^* = -\frac{\sum_{t=1}^{T} k_t}{T}$ and $e_l^* = cov(g_{lt}, k_t)/var(g_{lt})$ for $l = 1, \dots, L$.
- $\theta_0$ can be transformed back into the original $\theta$ by putting:
  - $d^0 = \frac{1}{d^*}$, then $\tilde{b}_x = \frac{b_x}{d^0} \Rightarrow b_x = \frac{\tilde{b}_x}{d^*}$;
  - $c^0 = -c^* d^*$, then $\tilde{a}_x = a_x - b_x c^0 \Rightarrow a_x = \tilde{a}_x - \frac{\tilde{b}_x}{d^*} c^* d^* = \widetilde{a_x} - \tilde{b}_x c^*$;
  - $e^0 = -e^* d^*$, then $\tilde{c}_{l,x} = c_{l,x} + e^0 b_x = c_{l,x} - e^* \tilde{b}_x \Rightarrow c_{l,x} = \tilde{c}_{l,x} + e^* \tilde{b}_x$;
  - It follows that $\widetilde{k_t} = d^0 \left(k_t - \sum_l e^0 g_{l,t} + c^0\right) = \frac{1}{d^*} \left(k_t + d^* \sum_l e^* g_{l,t} - c^* d^*\right) \Rightarrow k_t = d^* \left(\widetilde{k_t} - \sum_l e^* g_{l,t} + c^*\right)$. The parametrization is invariant to $c^*$, $d^*$ and $e^*$.
- Consider $\theta \neq \tilde{\theta}$.
  1. If $a_x \neq \tilde{a}_x$ for some $x$, then
  $$\frac{1}{T} \sum_{t=1}^{T} \mu_{x,t} = a_x \neq \tilde{a}_x = \frac{1}{T} \sum_{t=1}^{T} \tilde{\mu}_{x,t},$$
  since $k_t$ and $g_{l,t}$ sum to 0 (the singular value decomposition is performed on the scaled matrix of the external variables).
  2. If $h_{l,x} \neq \tilde{h}_{l,x}$ for some $l$ and $x$, since all $g_{l,t}$ are uncorrelated with each other, with $k_t$ and sum to 0 over $t$, then

$$\sum_{t=1}^{T} g_{l,t}\mu_{x,t} = h_{l,x}\sum_{t=1}^{T} {g^2}_{l'}, t \neq \tilde{h}_{l,x}\sum_{t=1}^{T} {g^2}_{l'}, t = \sum_{t=1}^{T} g_{l,t}\tilde{\mu}_{x,t}.$$

3. If $a_x = \tilde{a}_x$ and $h_{l,x} = \tilde{h}_{l,x}$ for all $x$ and $l$, but $k_t \neq \tilde{k}_t$ for some $t$, since $\sum_{x=1}^{N} b_x = 1$, it holds that

$$\begin{aligned}\sum_{x=1}^{N} \mu_{x,t} &= k_t + \sum_{x=1}^{N} a_x + \sum_{l=1}^{L}\sum_{t=1}^{T} h_{l,x} g_{l,t} \\ &\neq \tilde{k}_t + \sum_{x=1}^{N} \tilde{a}_x + \sum_{l=1}^{L}\sum_{t=1}^{T} \tilde{h}_{l,x} g_{l,t} \\ &= \sum_{x=1}^{N} \tilde{\mu}_{x,t}\,.\end{aligned}$$

4. If $a_x = \tilde{a}_x$, $h_{l,x} = \tilde{h}_{l,x}$ for all $x$ and $l$, $k_t = \tilde{k}_t$ for all $t$, but $b_x \neq \tilde{b}_x$ for some $x$, since $k \neq 0$, there exists some $t$ for which $k_t = \tilde{k}_t \neq 0$. Then $\sum_{t=1}^{T} k_t^2 = \sum_{t=1}^{T} \tilde{k}_t^2$. However, since $k_t$ sum to 1 and are uncorrelated with $g_{l,t}$,

$$\begin{aligned}\sum_{t=1}^{T} k_t\mu_{x,t} &= b_x \sum_{t=1}^{T} k_t^2 \\ &\neq \tilde{b}_x \sum_{t=1}^{T} \tilde{k}_t^2 \\ &= \sum_{t=1}^{T} \tilde{k}_t\tilde{\mu}_{x,t}\,.\end{aligned}$$

## 3.3 Variables

The external variables chosen are meant to represent widely available and easily measurable variables that may have a plausible, although perhaps weak or indirect, effect on mortality. This is consistent with the overall focus on forecast improvement and scenario building. They are as follows (sources in parentheses):

- average height of men aged 18 ((NCD-RisC) (2016) and (NCD-RisC) (2020) estimate, considered as non-stochastic quantities)
- real GDP per capita (Feenstra et al. (2015) and PWT)
- age-standardized share of men with raised blood pressure (Zhou et al. (2017), NCD-RisC estimate, considered as non-stochastic quantities)
- fruit consumption per capita[1] (FAO)

[1] Both fruit consumption and vegetable consumption data are sourced from FAO's food balances survey (FBS). Since FBS underwent a change in methodology starting in 2014, there is a break in the series, which can be substantial. The old methodology and new methodology series have been reconciled by multiplying the new methodology (post-2014) series by a coefficient calculated as the ratio of average consumption in 2010-2013 and average consumption in 2014-2017.

- vegetable consumption per capita (FAO)
- daily supply of calories per person (FAO)
- recorded alcohol consumption in liters per capita (15+) (WHO and Wine Economics Research Centre, University of Adelaide, Holmes and Anderson (2017))
- cigarette consumption (International Cigarette Consumption Database, Poirier et al. (2019))
- surface temperature anomaly in degrees Celsius (difference between average country temperature and 1961-1990 global average temperature) (HadCRUT4)
- fossil fuel consumption per capita (BP Statistical Review of World Energy via Our World in Data)

The preceding variables have both data available for a very high share of HMD countries and exhibit a high correlation with both age-specific death rates and the $k_t$ time index of the Lee-Carter model.

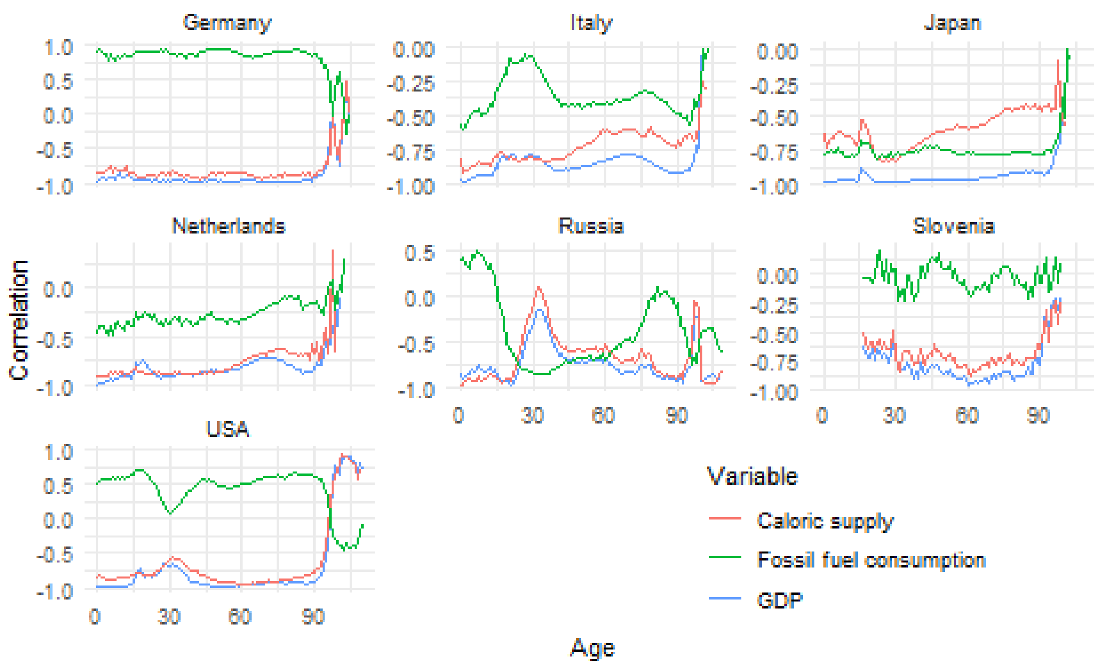


*Figure 3.1: Correlations between mortality and three potential covariates, by age*

The graphs of correlations with mortality by age and country show distinctive patterns, displaying either correlations above 0.75 in absolute value for most ages or smaller correlations that are erratic across ages. An example with three variables in seven countries is shown in figure 3.1. Correlations usually become much weaker or disappear entirely at ages above 90. In some cases, e.g. for the share of men that are obese, the correlation with mortality weakens markedly for ages from 20 to 40, around the mortality hump. In countries with a more complicated mortality history like Russia or with shorter time series like Slovenia, correlations for a given variable across ages are more likely to be erratic than in countries with longer time series and a simpler mortality history.

The following table summarizes the correlation between yearly age-standardized mortality rates and external factors for 7 HMD countries. The age weights used are the WHO 2000-2025 Standard Million (Ahmad et al., 2001).

*Table 3.1: Correlations between external variables and age-standardized mortality rates*

| Variable | Italy | Slovenia | Russia | Netherlands | Germany | Japan | USA |
|---|---|---|---|---|---|---|---|
| Real GDP per capita | -0.977 | -0.946 | -0.835 | -0.967 | -0.993 | -0.954 | -0.984 |
| Temperature anomalies | -0.807 | -0.330 | -0.273 | -0.778 | -0.274 | -0.453 | -0.570 |
| Fossil fuel consumption per capita | -0.489 | 0.321 | -0.681 | -0.278 | 0.932 | -0.804 | 0.609 |
| Caloric supply per capita | -0.612 | -0.819 | -0.686 | -0.439 | -0.869 | 0.022 | -0.899 |
| Share of men with raised blood pressure | 0.945 | 0.990 | 0.681 | 0.989 | 0.994 | 0.989 | 0.962 |
| Average height of men | -0.978 | -0.978 | -0.697 | -0.743 | -0.988 | -0.753 | 0.077 |
| Fruit and vegetable consumption per capita | -0.379 | 0.114 | -0.630 | -0.621 | 0.102 | 0.900 | -0.416 |
| Alcohol consumption liters per capita | 0.959 | 0.578 | 0.285 | 0.890 | 0.977 | -0.774 | 0.758 |
| Cigarette consumption per capita | 0.796 | 0.006 | -0.093 | 0.836 | 0.879 | 0.834 | 0.987 |

## 3.4 Stationarity of mortality time indices

Having covariates that are correlated with mortality is not, by itself, enough to build a mortality model, since if any time series is a non-stationary process, the model estimates will be inconsistent and the correlations may be spurious. It is therefore necessary to investigate whether mortality rates - summarized by the time index $k_t$ of the Lee-Carter model - and external variables are stationary or not. In the latter case, cointegration analysis needs to be performed in order to ascertain whether the time series have a common stochastic trend.

As in Seklecka et al. (2019), the Lee-Carter $k_t$ index for male mortality and all countries has been tested for stationarity with both the Phillips-Perron (Phillips and Perron, 1988) and KPSS (Kwiatkowski et al., 1992) tests, for which the null hypotheses are non-stationarity and stationarity, respectively. The version of the KPSS test for which the null is trend stationarity has been used.

The two tests agree on non-stationarity of $k_t$ (by not rejecting the null hypothesis for the Phillips-Perron test and by rejecting it for the KPSS test, in both cases at the 5% level) in 35 cases out of 42 countries. The tests disagree on the non-stationarity of $k_t$ for Chile, Croatia, Japan, Republic of Korea and Russia. The $k_t$ time index is stationary for Hong Kong and Taiwan. The detailed results are presented in table 3.2, with p-values>0.1 for the KPSS test presented as 0.1 and p-values<0.01 presented as 0.01.

*Table 3.2: Phillips-Perron and KPSS tests on the Lee-Carter time index*

| **Country** | **PP statistic** | **PP p-value** | **KPSS statistic** | **KPSS p-value** |
|---|---|---|---|---|
| Australia | -2.013 | 0.570 | 0.422 | 0.010 |
| Austria | -1.371 | 0.831 | 0.444 | 0.010 |
| Belgium | -1.801 | 0.656 | 0.435 | 0.010 |
| Bulgaria | -2.581 | 0.339 | 0.216 | 0.010 |
| Belarus | -0.721 | 0.963 | 0.244 | 0.010 |
| Canada | -1.117 | 0.913 | 0.442 | 0.010 |
| Chile | -2.650 | 0.323 | 0.129 | 0.082 |
| Croatia | -3.539 | 0.059 | 0.131 | 0.078 |
| Hong Kong | -4.389 | 0.010 | 0.072 | 0.100 |
| Switzerland | -2.274 | 0.464 | 0.438 | 0.010 |
| Czechia | -0.271 | 0.989 | 0.443 | 0.010 |
| East Germany | -1.938 | 0.600 | 0.402 | 0.010 |
| West Germany | -2.262 | 0.469 | 0.370 | 0.010 |
| Denmark | -0.357 | 0.986 | 0.454 | 0.010 |
| Spain | -3.269 | 0.084 | 0.292 | 0.010 |
| Estonia | 0.038 | 0.990 | 0.348 | 0.010 |
| Finland | -1.314 | 0.855 | 0.417 | 0.010 |
| France | -1.870 | 0.628 | 0.439 | 0.010 |
| Greece | -2.852 | 0.240 | 0.261 | 0.010 |
| Hungary | -1.014 | 0.929 | 0.373 | 0.010 |

| Country | PP statistic | PP p-value | KPSS statistic | KPSS p-value |
|---|---|---|---|---|
| Ireland | -0.333 | 0.987 | 0.450 | 0.010 |
| Iceland | -2.819 | 0.242 | 0.390 | 0.010 |
| Israel | -3.030 | 0.174 | 0.169 | 0.031 |
| Italy | -1.511 | 0.774 | 0.459 | 0.010 |
| Japan | -3.770 | 0.025 | 0.383 | 0.010 |
| Republic of Korea | -3.101 | 0.153 | 0.094 | 0.100 |
| Lithuania | -2.557 | 0.349 | 0.263 | 0.010 |
| Luxembourg | -2.899 | 0.211 | 0.360 | 0.010 |
| Latvia | -0.309 | 0.987 | 0.330 | 0.010 |
| Netherlands | 0.046 | 0.990 | 0.441 | 0.010 |
| Norway | 0.630 | 0.990 | 0.458 | 0.010 |
| New Zealand | -1.652 | 0.717 | 0.461 | 0.010 |
| Poland | -0.374 | 0.984 | 0.376 | 0.010 |
| Portugal | -2.209 | 0.490 | 0.400 | 0.010 |
| Russia | -1.301 | 0.856 | 0.135 | 0.070 |
| Slovakia | 0.676 | 0.990 | 0.368 | 0.010 |
| Slovenia | -1.754 | 0.669 | 0.211 | 0.012 |
| Sweden | -0.869 | 0.951 | 0.469 | 0.010 |
| Taiwan | -3.594 | 0.043 | 0.132 | 0.075 |
| Ukraine | -1.405 | 0.814 | 0.152 | 0.045 |
| United Kingdom | -1.927 | 0.605 | 0.436 | 0.010 |
| USA | -1.365 | 0.834 | 0.284 | 0.010 |

## 3.5 Structural breaks in mortality and covariates

Following Boonen and Li (2017) and Berkum et al. (2016), the possible presence of structural breaks has been investigated. While in Boonen and Li (2017) structural breaks are only used to calibrate forecasts, a non-stationary series may instead be trend stationary with structural breaks (Perron, 1989), which may alleviate issues with variables who show a high order of integration. Moreover, the presence of a common structural break in both mortality series and an external covariate, with a similar trend for the two variables both before and after the break, would reinforce the credibility of a link between them.

The variables considered were the $k_t$ Lee-Carter mortality index and the stochastic covariates[2]: logarithm of real GDP, temperature anomalies, fossil fuel consumption, caloric supply, fruit and vegetable consumption, alcohol consumption, cigarette consumption.

The methods to test for structural change are the generalized fluctuation tests, described by Zeileis et al. (2010) and implemented in the R package `strucchange`. Recursive residuals have been used to analyze both cumulative sums of residuals (CUSUM processes) and moving sums of residuals (MOSUM processes). For each time series, a linear time trend is estimated and then residuals are calculated. The series is deemed to have at least a structural break if the null hypothesis of no break is rejected by the structural change test with $\alpha = 0.05$ for at least one of the CUSUM and MOSUM processes. The optimal break points are then estimated by using the algorithm of Bai and Perron (2003), with up to 1 break point identified per series.

Out of 42 countries, 35 have at least one break point in either the $k_t$ Lee-Carter time index or in a covariate. Belgium, Hong Kong, Israel, Republic of Korea, Luxembourg, New Zealand and Taiwan show no breakpoints. More specifically, 21 countries have at least one structural break in the $k_t$ index. Temperature anomalies show a structural break in 3 countries, fossil fuel consumption in 11 countries, fruit and vegetable consumption in 13 countries, while the other covariates have a structural break in between 16 to 20 countries.

For each of the 21 structural breaks in the time index $k_t$, it has been checked whether a covariate had a structural break in the same or in the preceding year. Fossil fuel consumption has a concurrent or immediately preceding break point in 2 cases, same as cigarette consumption per capita, followed by logarithm of real GDP, caloric supply, alcohol consumption and fruit and vegetable consumption with one case.

To sum up, both the mortality index $k_t$ and the external variables are often non-stationary and prone to having structural breaks even in relatively short time series (40 years or less). It does not appear, though, that breaks in $k_t$ are systematically preceded or followed by breaks in one or more of the external variables, hence there is no evidence that a structural break in a variable causes a break in $k_t$ or vice versa.

Additionally, to check whether the relationship between mortality and the other variables is stable over time, it has been investigated whether there is a structural break in the series of the residuals of the time index $k_t$ regressed on the stochastic external variables. The only country to show such a break is East Germany, with a break estimated in 1996. When regressed on all variables, including the non-stochastic share of men with raised blood pressure and male height at age 18, Spain shows a structural break around 1985 and Poland in 1988. Except these three countries, the relationship between te mortality index and the external variables is stable over time.

---

[2] As mentioned in section 3.3, average height at age 18 and share of men with raised blood pressure are modeled quantities and are therefore considered non-stochastic.

## 3.6 Integration of covariates

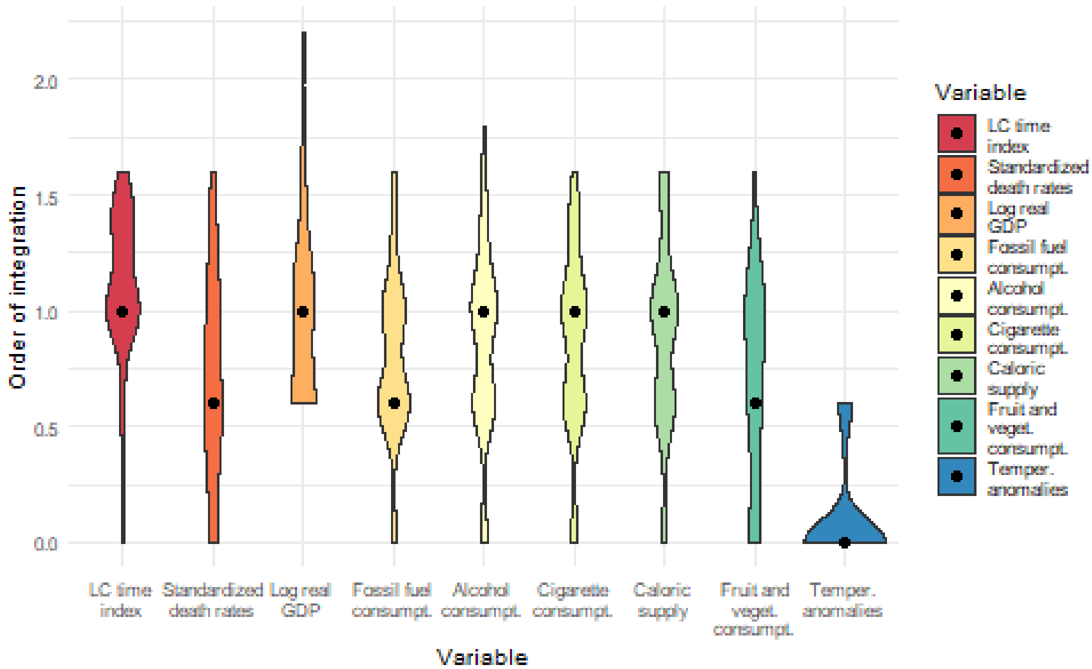


*Figure 3.2: Orders of integration for all countries by variable*

It has been shown that both the time index $k_t$ and the external variables are non-stationary in most cases. In order to be able to model their long-term relationship, non-stationary variables need to be cointegrated and a prerequisite for cointegration is that the series are I(1). To investigate whether this is the case, the order of integration of all variables for all countries has been explored with both the Phillips-Perron and the KPSS tests (both at $\alpha = 0.05$), with the following procedure:

- Run both the Phillips-Perron test and the KPSS test on the series;
- If both tests agree on non-stationarity (that is, the Phillips-Perron does not reject the null and the KPSS test rejects the null), differentiate the series and repeat the procedure on the differentiated series;
- If both tests disagree with each other (i.e. they both fail to reject the null), for illustrative purposes only add 0.6 to the order of integration if the Phillips-Perron test points to non-stationarity and 0.4 if the KPSS test points at non-stationarity.

The results are presented in figure 3.2. Temperature anomalies are mostly stationary, with 5 countries non-stationary according to the Phillips-Perron test only and 1 country non-stationary according to the KPSS test only. The other variables are generally non-stationary, with multiple cases of discordance between the PP and KPSS tests about whether a series is I(1) or I(2). Alcohol consumption in the Republic of Korea, which has a short time series, is I(3) according to the Phillips-Perron test and I(0) according to the KPSS test, while for real GDP in Croatia is I(3) for the Phillips-Perron test and I(1) for the KPSS test.  The procedure has been repeated considering structural breaks as well for series that had an order of integration of 1 or above. For series that had an order of integration below 1, structural breaks were ignored. Due to the shortness of the series, considering the structural break usually results in the null not being rejected for both the Phillips-Perron and the KPSS tests.

All in all, temperature anomalies are mostly stationary and the other variables are non-stationary. No series is unambiguously I(2) or higher, though for a sizable number of series the tests disagree whether they are I(1) or I(2). When considering structural breaks, the series are either I(1) or are non-stationary according to the Phillips-Perron test and stationary according to the KPSS test. Therefore it cannot be argued that the series are stationary with structural breaks.

## 3.7 Cointegration analysis

As already outlined, most of the time series considered are I(1). Following the methodology outlined in Seklecka et al. (2019), the presence of cointegration relationships in the data has been explored. The procedure by Johansen (1991) has been applied to all countries with at least 20 years of data (27 countries), assuming a linear trend in cointegration. The results are presented in table 3.3. With the exception of Denmark and Finland, all countries have at least one cointegration relationship with a p-value under 5% and 23 countries (85%) have at least one cointegration relationship with a p-value under 1%. In both Denmark and Finland the p-value is just above 5%.

*Table 3.3: Cointegration relationships between the Lee-Carter time index and external variables*

| Country | Number of cointegration relationships at 99% significance | Number of cointegration relationships at 95% significance |
|---|---|---|
| Australia | 1 | 3 |
| Austria | 1 | 1 |
| Bulgaria | 0 | 2 |
| Canada | 1 | 2 |
| Chile | 6 | 7 |
| Switzerland | 1 | 1 |
| Czechia | 7 | 8 |
| East Germany | 1 | 1 |
| West Germany | 1 | 1 |
| Denmark | 0 | 0 |
| Spain | 2 | 3 |
| Finland | 0 | 0 |
| France | 1 | 1 |
| Greece | 2 | 3 |

| Country | Number of cointegration relationships at 99% significance | Number of cointegration relationships at 95% significance |
|---|---:|---:|
| Hungary | 2 | 3 |
| Ireland | 1 | 1 |
| Iceland | 1 | 3 |
| Italy | 2 | 3 |
| Japan | 2 | 2 |
| Netherlands | 2 | 3 |
| Norway | 2 | 3 |
| Poland | 0 | 3 |
| Portugal | 1 | 1 |
| Slovakia | 7 | 7 |
| Sweden | 1 | 2 |
| United Kingdom | 1 | 1 |
| USA | 3 | 3 |

# 4 Results

The SEV model is applied to male mortality between the ages of 40 to 90, with a maximum of two principal components and a threshold of 15% of variance explained by the principal component in order for it to be included in the model: a brief discussion of alternative specifications is presented in section 4.4. The overall performance of the SEV model for all countries analyzed is discussed first, using several goodness of fit measures comprising both absolute and percentage errors. The distribution of age at death is concentrated on higher ages, therefore absolute residuals will be more influenced by the fit at old ages, while percentage residuals are more sensitive at younger ages, when the number of deaths is small and a small residual in absolute terms can lead to large errors in percentage terms. Subsequently, forecasting performance is assessed for a variety of jump-off years. Finally, the results for a few countries are discussed individually, with a specific focus on whether the model fit captures the mortality trend adequately by analyzing the model residuals, again both in raw number of deaths and in percentage terms.

## 4.1 Goodness of fit and model comparison

The fit of the SEV models is compared to the Lee-Carter model and to the Niu-Melenberg model, estimated on the same set of years and ages, using the Mean Absolute Deviation (MAD), Mean

Absolute Percentage Error (MAPE) and the Bayesian Information Criterion (BIC). They are defined as follows, with the predicted deaths $\widehat{D}_{x,t} = exp\{a_x + \sum_{j=1}^{J} b_{j,x}k_{j,t} + \sum_{l=1}^{L} c_{l,x}d_{l,x}g_{l,t}\}E_{x,t}$:

$$MAD = \frac{\sum_x \sum_t \left|D_{x,t} - \widehat{D}_{x,t}\right|}{XT}$$

$$MAPE = \frac{\sum_x \sum_t \frac{\left|D_{x,t} - \widehat{D}_{x,t}\right|}{D_{x,t}}}{XT}$$

$$BIC = m\ logM - 2log\hat{L},$$

with $X$ being the number of ages, $T$ the number of years, $m$ the number of parameters, $M$ the number of observations and $\hat{L}$ the likelihood in the model. A lower value of the BIC indicates a better fit.

The results are presented in figure 4.1 and 4.1. The SEV models outperform both the Lee-Carter and the Niu-Melenberg models in terms of both MAD and MAPE: since the Niu-Melenberg fits better than the Lee-Carter, only the comparison with the Niu-Melenberg is shown. In terms of BIC, the SEV models has either a lower BIC or a very close BIC to both the Lee-Carter and Niu-Melenberg models, outperforming the Lee-Carter model for 18 countries out of 37 and the Niu-Melenberg model for 12 countries out of 37. The proposed single-population models therefore improve the fit beyond their cost in additional model complexity for between a third and half of the countries examined, while in the other countries the improved fit comes at a negligible added complexity.

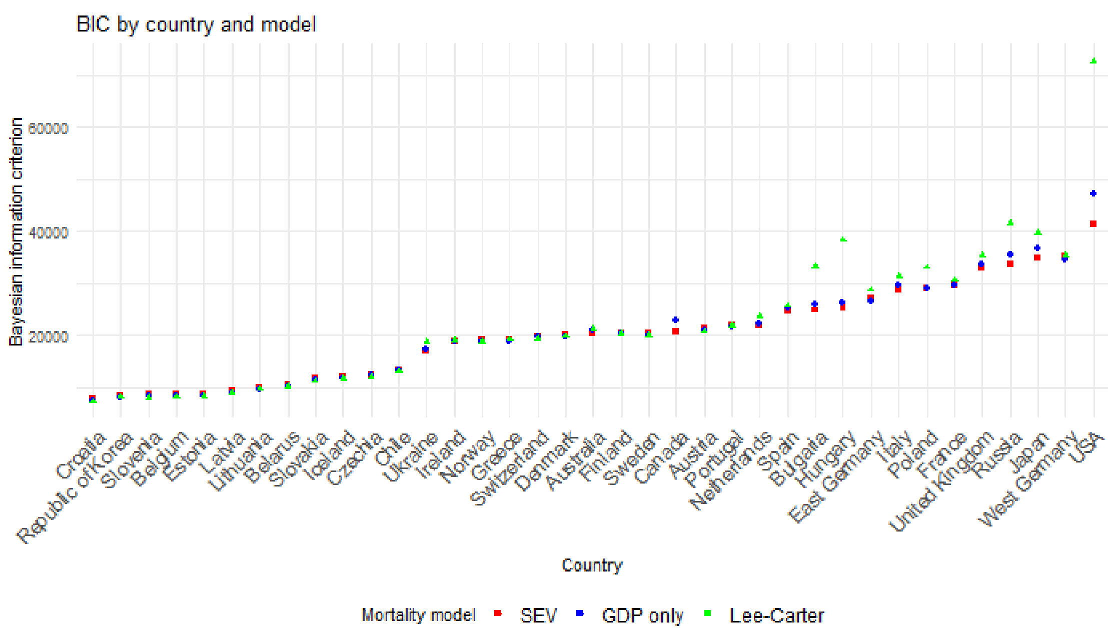


*Figure 4.1: Bayesian Information Criterion (BIC) for SEV, Niu-Melenberg and Lee-Carter models*

*Table 4.1: Mean absolute deviation (MAD), mean absolute percentage error (MAPE), Bayesian Information Criterion (BIC) and number of principal components for single-population models and*

*equivalent Niu-Melenberg models. Countries where the single-population model outperforms the Lee-Carter model are in bold.*

| Country | MAD | MAPE | BIC | N. PC | MAD Niu-Melenberg | MAPE Niu-Melenberg | BIC Niu-Melenberg |
|---|---|---|---|---|---|---|---|
| **Australia** | **28.793** | **3.537** | **20,293** | **2** | **34.825** | **4.108** | **20,934** |
| Austria | 23.353 | 4.190 | 21,217 | 2 | 23.140 | 4.194 | 20,821 |
| Belarus | 34.242 | 3.053 | 10,543 | 2 | 35.036 | 3.117 | 10,269 |
| Belgium | 20.751 | 3.026 | 8,544 | 2 | 22.746 | 3.262 | 8,320 |
| **Bulgaria** | **38.944** | **4.096** | **24,816** | **2** | **43.727** | **4.527** | **25,750** |
| **Canada** | **36.700** | **2.629** | **20,615** | **2** | **51.607** | **3.761** | **22,687** |
| Chile | 28.053 | 3.993 | 13,360 | 2 | 29.247 | 4.128 | 13,121 |
| Croatia | 16.498 | 4.443 | 7,669 | 2 | 17.186 | 4.510 | 7,407 |
| Czechia | 26.058 | 3.153 | 12,436 | 2 | 27.136 | 3.270 | 12,217 |
| Denmark | 18.213 | 4.584 | 20,102 | 2 | 18.484 | 4.725 | 19,801 |
| East Germany | 52.937 | 4.500 | 27,072 | 2 | 53.026 | 4.422 | 26,614 |
| Estonia | 8.477 | 6.237 | 8,791 | 2 | 8.864 | 6.491 | 8,501 |
| Finland | 18.537 | 5.287 | 20,434 | 2 | 19.997 | 5.660 | 20,404 |
| France | 93.696 | 2.602 | 29,613 | 2 | 94.036 | 2.715 | 29,456 |
| Greece | 29.116 | 3.995 | 19,020 | 2 | 29.944 | 4.130 | 18,772 |
| **Hungary** | **44.644** | **3.868** | **25,360** | **2** | **50.506** | **4.298** | **26,233** |
| Iceland | 2.557 | Inf | 12,142 | 2 | 2.617 | Inf | 11,845 |
| Ireland | 13.309 | 6.077 | 18,763 | 2 | 13.956 | 6.688 | 18,687 |
| **Italy** | **97.001** | **2.400** | **28,763** | **2** | **108.931** | **2.644** | **29,705** |
| **Japan** | **176.346** | **2.493** | **34,846** | **2** | **189.605** | **2.836** | **36,546** |
| Latvia | 10.934 | 4.533 | 9,363 | 2 | 11.337 | 4.716 | 9,063 |
| Lithuania | 14.445 | 4.274 | 9,854 | 2 | 14.853 | 4.401 | 9,580 |
| **Netherlands** | **28.009** | **3.136** | **21,873** | **2** | **32.237** | **3.718** | **22,315** |
| Norway | 14.298 | 5.188 | 18,977 | 2 | 15.198 | 5.623 | 18,862 |
| Poland | 82.947 | 2.644 | 29,054 | 2 | 84.480 | 2.732 | 29,026 |

| Country | MAD | MAPE | BIC | N. PC | MAD Niu-Melenberg | MAPE Niu-Melenberg | BIC Niu-Melenberg |
|---|---|---|---|---|---|---|---|
| Portugal | 26.201 | 3.730 | 21,839 | 2 | 26.846 | 3.752 | 21,536 |
| Republic of Korea | 51.402 | 2.305 | 8,412 | 2 | 54.115 | 2.390 | 8,169 |
| **Russia** | **450.430** | **2.518** | **33,633** | **2** | **478.080** | **2.739** | **35,484** |
| Slovakia | 17.663 | 4.034 | 11,642 | 2 | 18.368 | 4.189 | 11,376 |
| Slovenia | 9.292 | 6.565 | 8,533 | 2 | 9.740 | 6.764 | 8,258 |
| **Spain** | **52.688** | **2.159** | **24,704** | **2** | **55.366** | **2.620** | **25,252** |
| Sweden | 20.364 | 3.639 | 20,451 | 2 | 20.758 | 3.786 | 20,165 |
| Switzerland | 17.157 | 4.396 | 19,892 | 2 | 17.666 | 4.519 | 19,635 |
| **Ukraine** | **154.880** | **2.591** | **17,090** | **2** | **158.623** | **2.670** | **17,130** |
| **United Kingdom** | **113.557** | **2.996** | **32,968** | **2** | **120.853** | **3.180** | **33,673** |
| **USA** | **276.788** | **1.819** | **41,230** | **2** | **318.170** | **2.205** | **46,994** |
| West Germany | 136.440 | 2.807 | 35,219 | 2 | 133.902 | 2.642 | 34,393 |

## 4.2 Forecasting performance

Even more important than fitting historical data is the model's ability to forecast future mortality rates. The forecasting performance of the single-population models is evaluated on historical data by estimating the model up to a jump-off year, forecasting the remaining years until the end of the sample and then comparing the forecasts with the actual mortality rates. The metric used is the relative root mean forecast square error (RMFSE), as in Boonen and Li (2017).

Given a jump-off year $\hat{u}$, a predicted logarithm of mortality rate $\widehat{log\ m_{x,\iota,t}}$ for country $i$, age $x$ and year $t$, and $U_i$ being the end of sample year for population $i$, the RMFSE is:

$$RMFSE_M(i,\hat{u}) = \sqrt{\frac{1}{N(U_i-\hat{u})}\sum_{u=\hat{u}+1}^{U_i}\sum_{x=0}^{N}\frac{\left(log\ m_{i,x,u} - \widehat{log\ m_{x,\iota,t}}\right)^2}{\left|log\ m_{i,x,u}\right|}}$$

The period term $k_t$ has been modeled as a random walk with drift. The models are evaluated with jump-off years between 2000 and 2010, in order to compare different forecast horizons. The results are presented for all countries in table 4.2 for years 2000, 2005 and 2010. For the whole span of jump-off years 2000-2010 the results are presented in figure 4.2 for a group of low-mortality countries with similar mortality paths and in figure 4.3 for a group of similar high-mortality countries.

The forecasting performance has been tested for a given jump-off year if the model had at least 10 years of data up to the jump-off year included, if data was available for at least one year after the jump-off year and if the country is big enough to have at least 5.000 deaths per year across all ages considered.

*Table 4.2: Ratio between root mean square forecast squared error (RMFSE) for single-population models and correspondent Niu-Melenberg (NM) models, jump-off year 2000, 2005 and 2010, only countries with at least 9 years of data up to jump-off year*

| Country | Model starting year | Ratio RMFSE/ RMFSE NM 2000 | Ratio RMFSE/ RMFSE NM 2005 | Ratio RMFSE/ RMFSE NM 2010 |
|---|---|---|---|---|
| Australia | 1975 | 1.069 | **0.685** | |
| Austria | 1975 | **0.758** | 1.147 | 1.454 |
| Belarus | 1997 | | | 1.188 |
| Belgium | 2000 | | | 1.376 |
| Bulgaria | 1975 | **0.758** | **0.752** | 1.192 |
| Canada | 1975 | 1.129 | **0.544** | **0.613** |
| Chile | 1992 | | **0.521** | **0.855** |
| Croatia | 2001 | | | **0.799** |
| Czechia | 1993 | | **0.849** | 1.299 |
| Denmark | 1975 | **1.000** | **0.919** | 1.145 |
| East Germany | 1975 | **0.856** | **0.981** | **0.998** |
| Estonia | 1996 | | 1.113 | **0.818** |
| Finland | 1975 | 1.006 | 1.044 | **0.820** |
| France | 1975 | **0.914** | 1.056 | 1.057 |
| Greece | 1981 | 1.187 | **0.992** | 1.172 |
| Hungary | 1975 | 1.050 | 1.607 | 1.006 |
| Ireland | 1975 | 1.044 | **0.775** | 1.163 |
| Italy | 1975 | **0.916** | 1.151 | 1.154 |
| Japan | 1975 | **0.964** | **0.919** | **0.830** |

| **Country** | **Model starting year** | **Ratio RMFSE/ RMFSE NM 2000** | **Ratio RMFSE/ RMFSE NM 2005** | **Ratio RMFSE/ RMFSE NM 2010** |
|---|---|---:|---:|---:|
| Latvia | 1996 | | **0.961** | **0.985** |
| Lithuania | 1996 | | 1.088 | 1.026 |
| Netherlands | 1975 | **0.953** | 1.026 | **0.891** |
| Norway | 1975 | 1.039 | 1.219 | **0.961** |
| Poland | 1975 | **0.939** | 1.476 | 1.113 |
| Portugal | 1975 | 1.520 | 2.217 | 1.509 |
| Russia | 1996 | | **0.531** | 1.493 |
| Slovakia | 1993 | | **0.893** | 1.216 |
| Slovenia | 1997 | | | **0.810** |
| Spain | 1975 | 1.257 | **0.862** | 1.933 |
| Sweden | 1975 | 1.114 | 1.070 | 1.584 |
| Switzerland | 1975 | **0.829** | **0.540** | 1.053 |
| Ukraine | 1996 | | **0.763** | **0.846** |
| United Kingdom | 1975 | **0.785** | **0.960** | 1.759 |
| USA | 1975 | 1.240 | 1.350 | 1.350 |
| West Germany | 1975 | **0.898** | **0.980** | 1.203 |

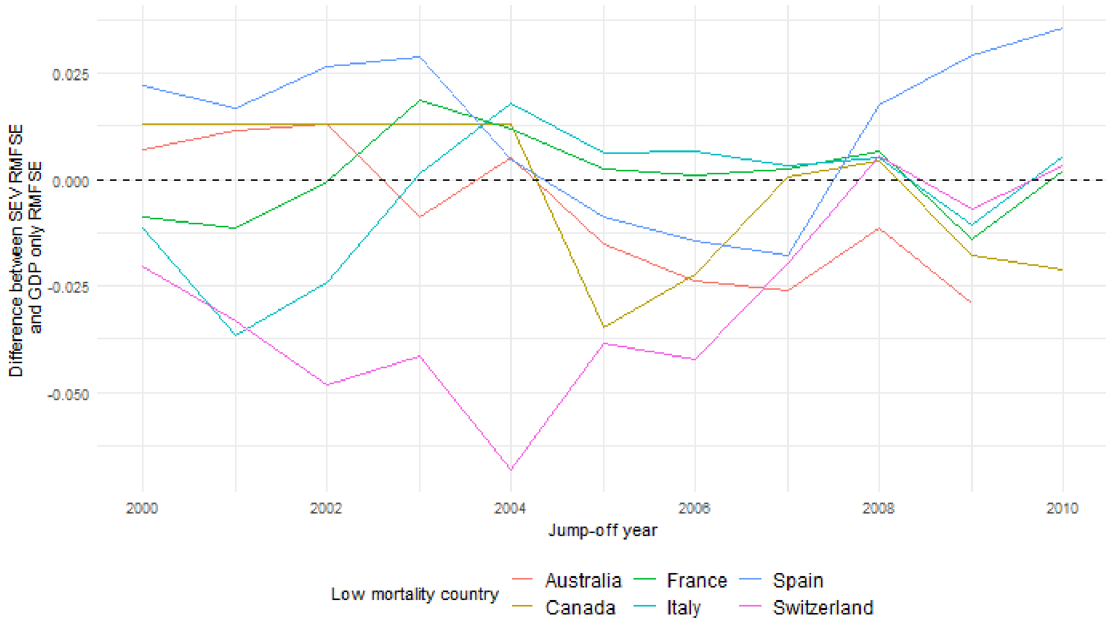


*Figure 4.2: Difference between RMFSE for the SEV model for for the Niu-Melenberg model, low mortality countries (SEV is performing better if the line is under the dashed line)*

For 12 countries out of 23 the ratio between the RMFSE of the SEV model and the Niu-Melenberg model is lower than 1, that is, the SEV model outperforms the Niu-Melenberg model in terms of RMFSE. For jump-off years 2005 and 2010 the countries outperforming the Niu-Melenberg model are 18 out of 31 and 12 out of 34, respectively. Overall, adding external variables beyond GDP improves the forecasting performance of the model for about half of the countries.

The better performing model can vary from year to year. As shown in figure 4.2, the SEV model is generally outperforming the Niu-Melenberg model for Australia, Switzerland and Canada[3], while the performance for France, Italy and Spain. Still, for all six countries there are years when the SEV model outperforms the Niu-Melenberg model and vice-versa.

[3] Covariate data is missing for Canada between 1999 and 2004, hence the last year with actual data is 1998 until 2005

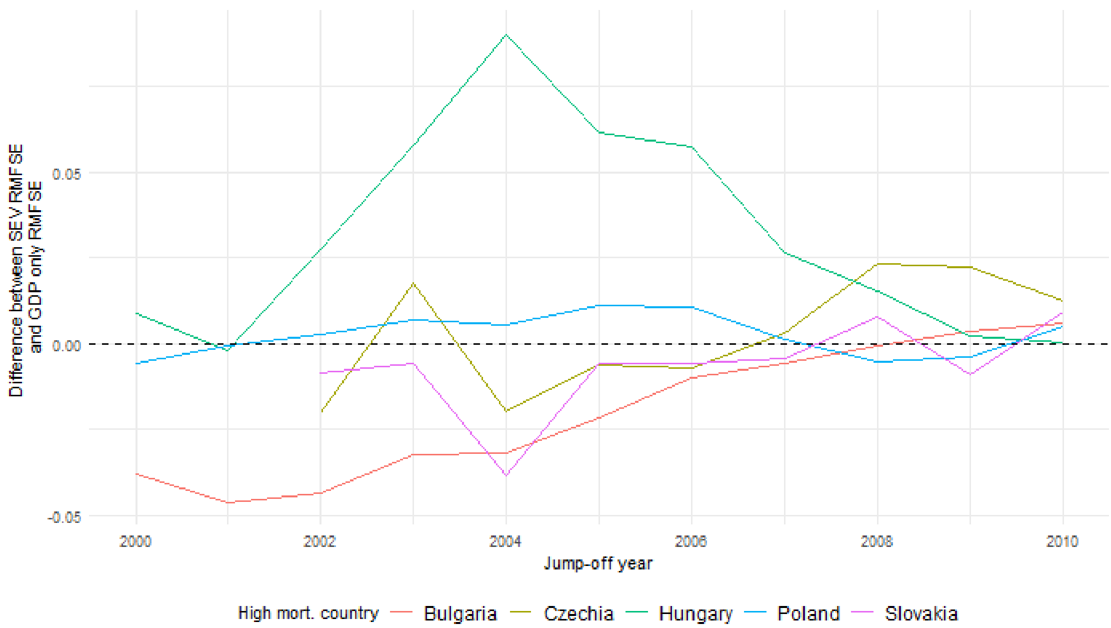


*Figure 4.3: Difference between RMFSE for the SEV model for for the Niu-Melenberg model, high mortality countries (SEV is performing better if the line is under the dashed line)*

This is even more evident in figure 4.3: the SEV model decidedly outperforms the Niu-Melenberg model in Bulgaria and, to a lesser extent, Slovakia, while it underperforms in Hungary. For all three countries, though, there are years where the opposite is true.

There is no clear relationship between the number of years used to train the model and the relative forecasting performance. The two countries with data starting in the 2000s (Belgium and Croatia) do not outperform the Niu-Melenberg model, but there are both countries with data starting in 1975 (the Netherlands) and countries with data from 1992 onwards (Russia) that outperform the Niu-Melenberg model for all applicable jump-off years.

## 4.3 Individual countries

### 4.3.1 Netherlands

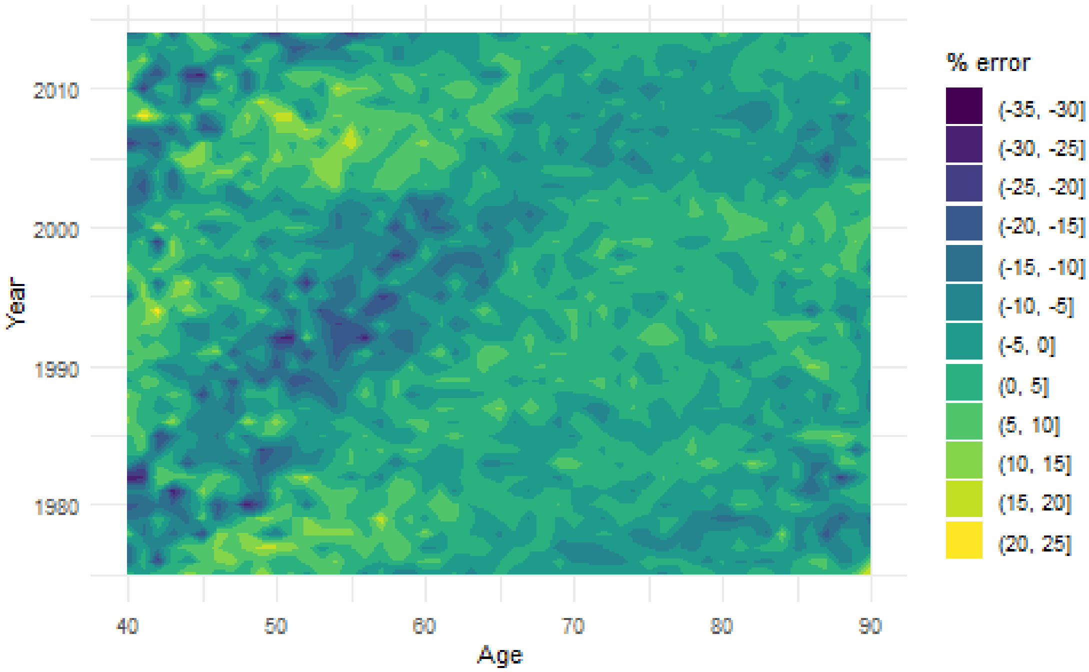


*Figure 4.4: Netherlands, actual minus predicted deaths as % of actual deaths by age and year*

Figure 4.4 shows the residuals by age and calendar year as a percentage of total deaths. There are no systematic age errors, which would show as vertical lines, or systematic period errors, which would show as horizontal lines. The errors vary at random and the largest percentage errors are in the 40-45 age range, where the absolute number of deaths is smaller and a small error in absolute terms is magnified in percentage terms. There are hints of a small cohort effect (diagonal bands) for the WW2 cohorts, which is mostly gone after year 2000.

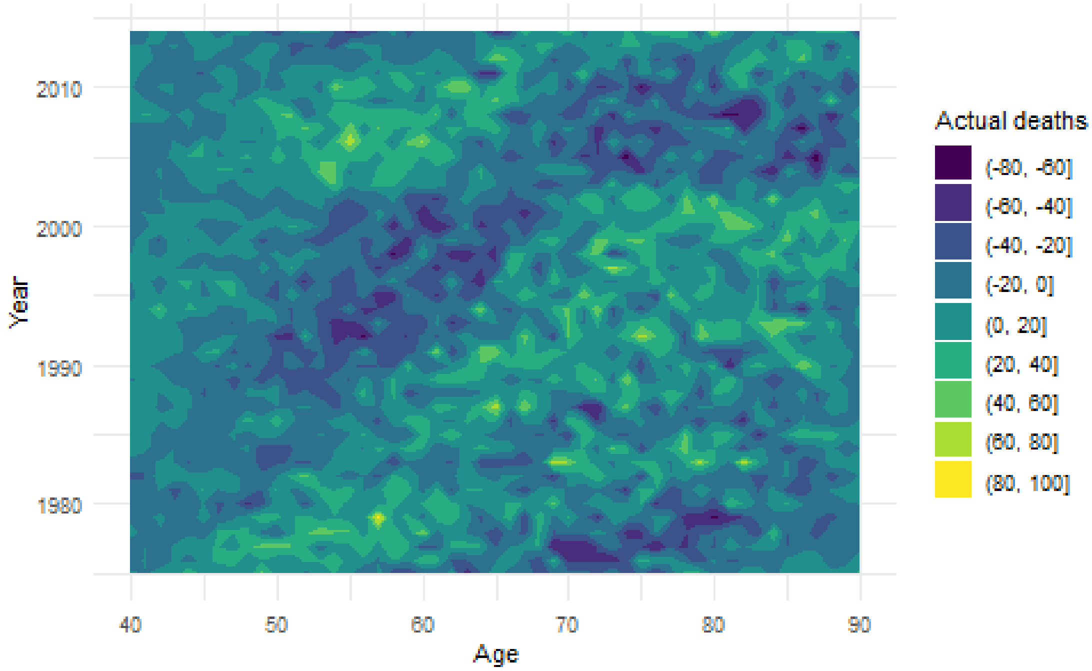


*Figure 4.5: Netherlands, actual minus predicted deaths by age and year*

The errors in absolute terms, shown in figure 4.5, show a similar pattern to the percentage errors. No significant structural patterns are visible and the occasional high residuals aren't clustered together.

### 4.3.2 Spain

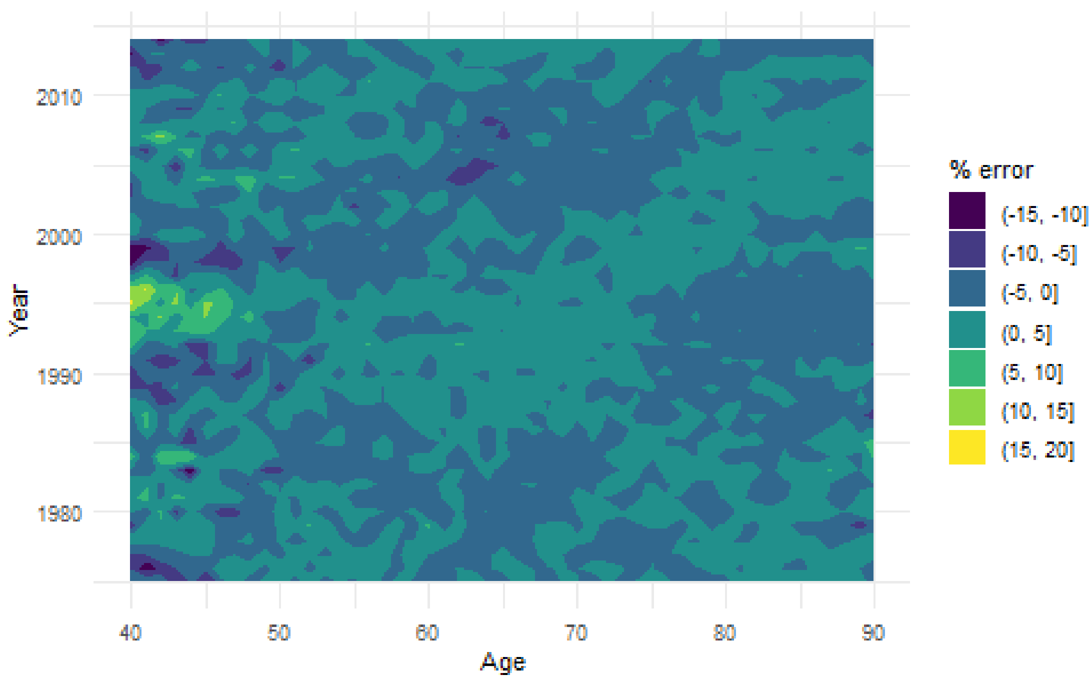


*Figure 4.6: Spain, actual minus predicted deaths as % of actual deaths by age and year*

For Spain, the percentage difference between the actual and fitted deaths is between -15% and 20%. The largest residuals in percentage terms are for the 40-50 age range around 1995. The timing and age range affected is consistent with the AIDS epidemic and the introduction of new therapies in the mid-90s: moreover, a version of the model comprising the age range 0-90 evidenced a spike in male deaths aged 30-40 in the same years. While the AIDS epidemic has not been explicitly modeled for comparability across countries, it's a component of mortality that could be straightforwardly added to the SEV model. There are no systematic age or period components visible. The residuals hint at a small cohort effects for those born during the Spanish Civil War.

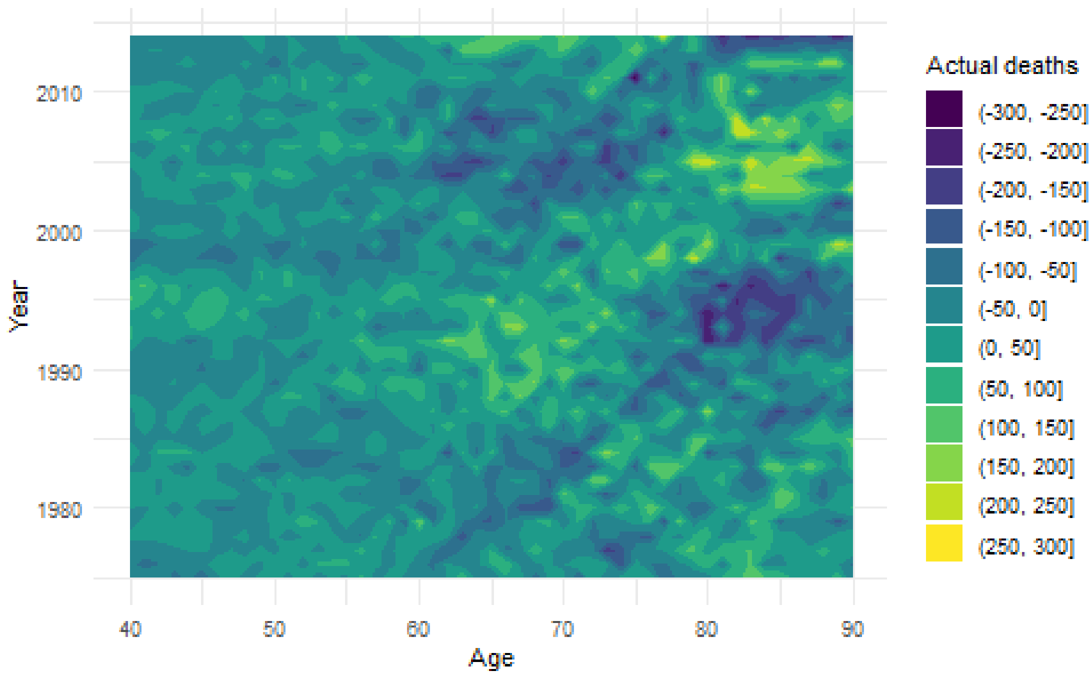


*Figure 4.7: Spain, actual minus predicted deaths by age and year*

The cohort effect due to the Spanish Civil war remains visible when residuals are calculated as difference in the number of deaths. The range of the residuals is [-300, 300], with the model predicting a higher number of deaths for ages 80+ in the mid-nineties.

### 4.3.3 Poland

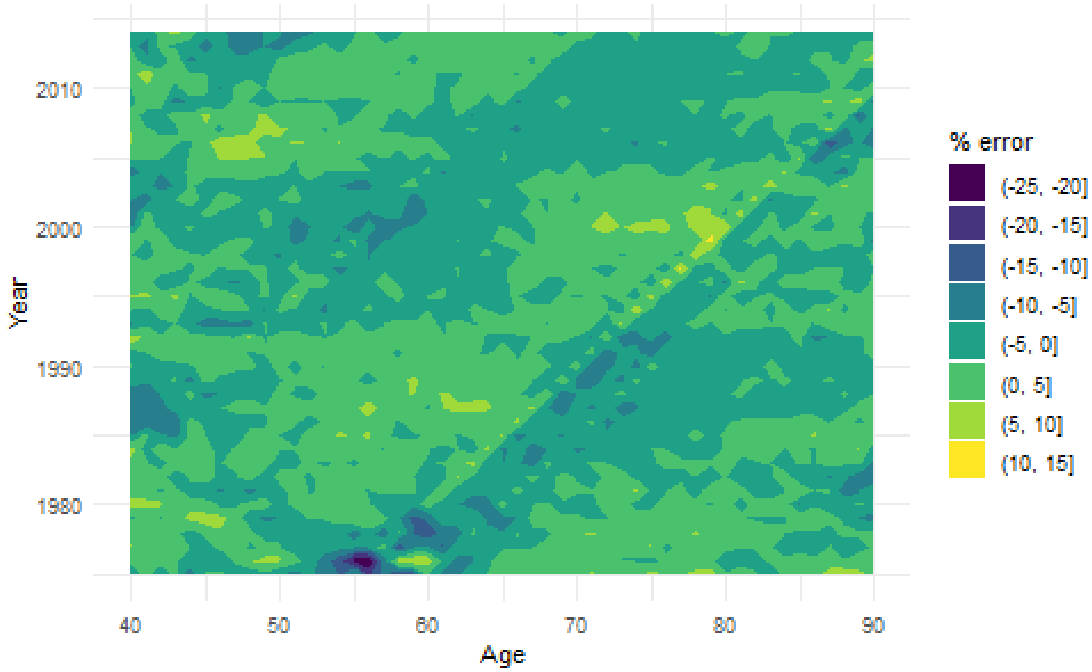


*Figure 4.8: Poland, actual minus predicted deaths as % of actual deaths by age and year*

The range of percentage residuals for Poland is from -25% to 15%, again comparable with the other countries. The residuals show two prominent cohort effects: one for the cohorts born around 1920 (Polish-Soviet war), a weaker one for the cohorts born during World War 1. No systematic age or period effect is visible, but the model predicts a larger number of deaths in 1975-77 around age 55, the cohorts for which the cohort effects are most prominent in the subsequent years.

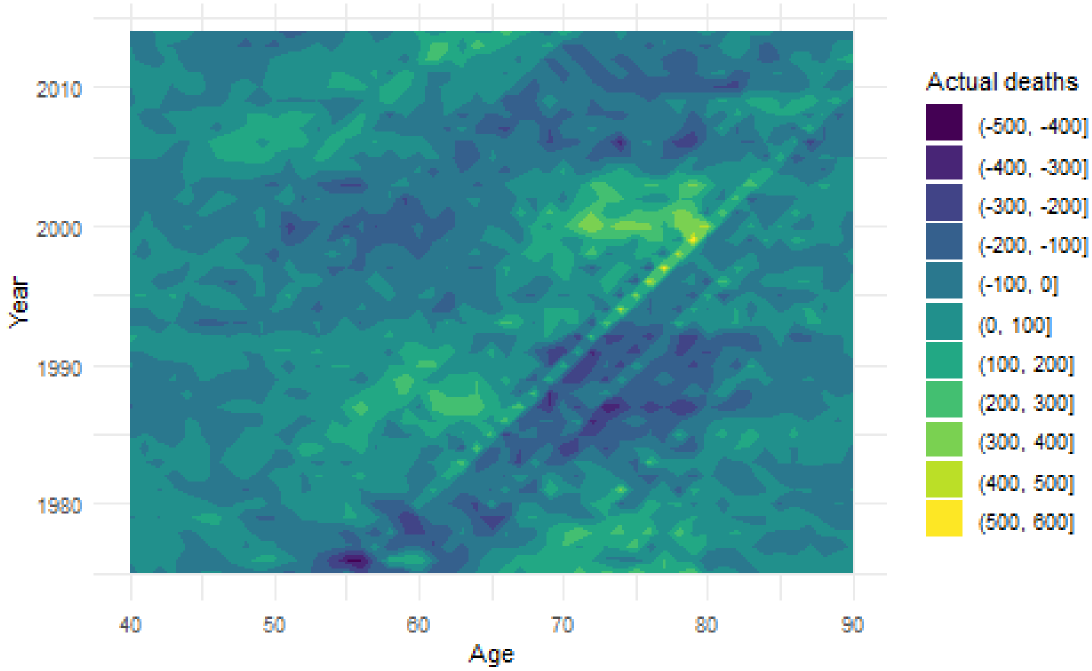


*Figure 4.9: Poland, actual minus predicted deaths by age and year*

The residual deaths range from -500 to 600 per year, with, again, the biggest errors related to the 1920 and WW1 cohorts. The graph also hints at a weak cohort effect for those born during WW2 and slightly overestimates deaths at ages 70 and higher in years 1988-1992, underestimating the deaths in the same age range around year 2000.

### 4.3.4 Sweden

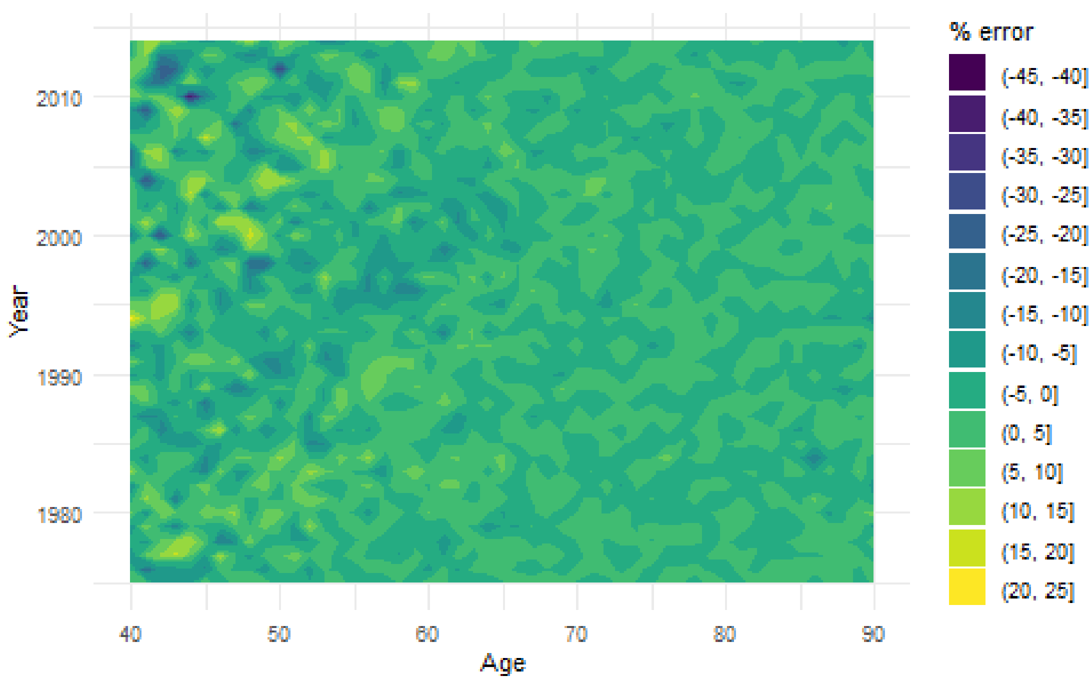


*Figure 4.10: Sweden, actual minus predicted deaths as % of actual deaths by age and year*

Sweden is the smallest country of the five considered individually and therefore the percentage residuals show the largest range, from -45% to 25%. The largest residuals in percentage occur for ages between 40 and 50 and the residuals appear entirely random, without any age, period or cohort effect.

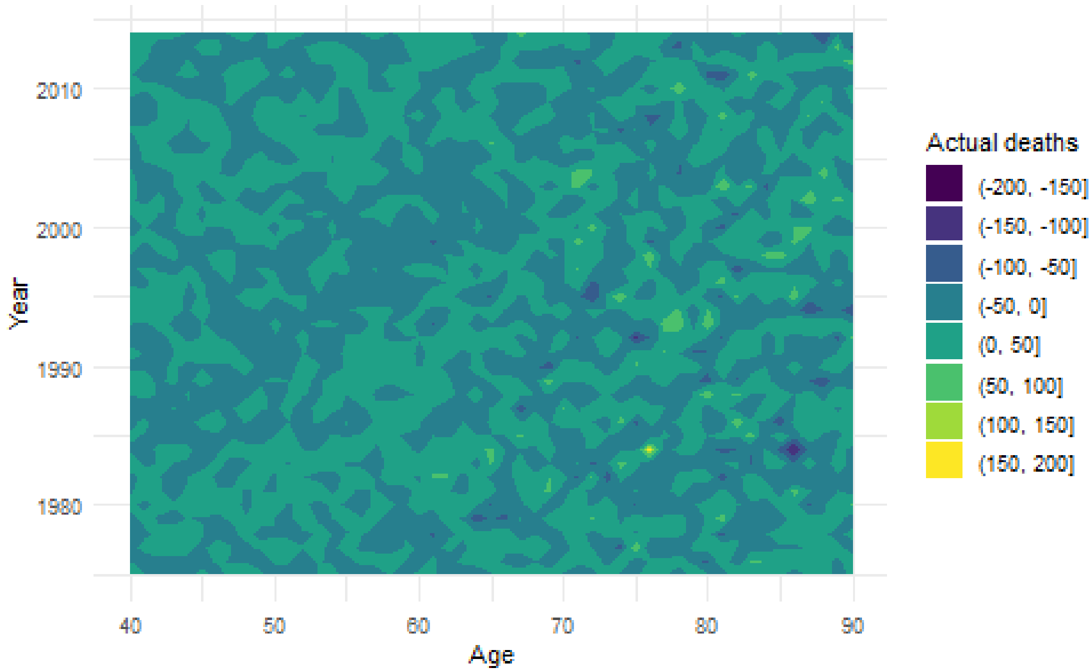


*Figure 4.11: Sweden, actual minus predicted deaths by age and year*

The absence of any pattern applies to the residual deaths as well. There are two large residuals in 1984 (170 for men aged 76 and -152 for men aged 86) and other smaller residuals scattered without a noticeable pattern.

### 4.3.5 USA

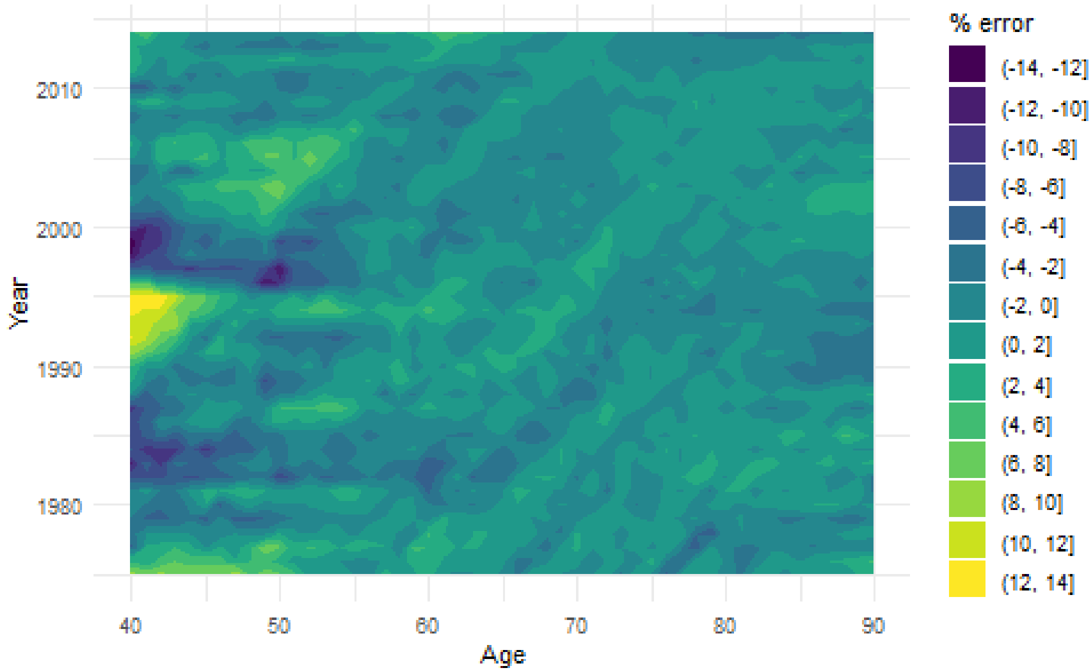


*Figure 4.12: USA, actual minus predicted deaths as % of actual deaths by age and year*

The United States are the largest country and hence the residuals in percentage terms are the smallest, ranging from -14% to 14%. Large positive residuals are present in the age range 40-50 up to 1995: as with Spain, a timing and age range consistent with the AIDS epidemic. There are several diagonal lines that hint at weak cohort effects.

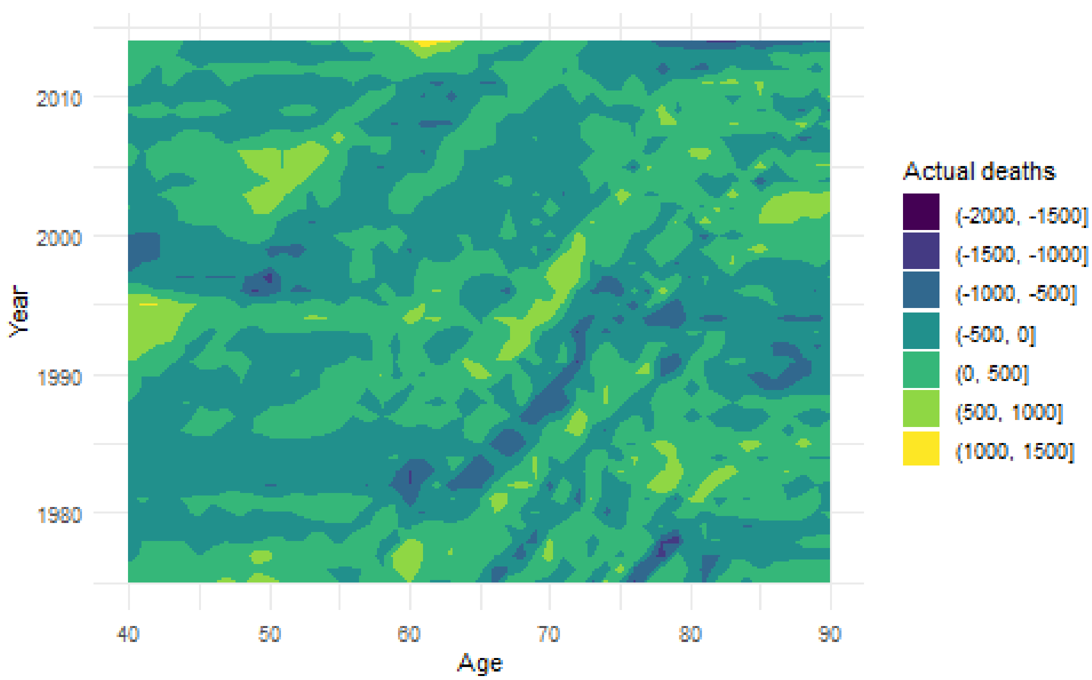


*Figure 4.13: USA, actual minus predicted deaths by age and year*

Examining the difference in actual and predicted deaths highlights the cohort effects, with the most prominent being the 1918-1919 cohorts. In 2014, the last year of data, the model noticeably underestimates deaths in the 60-65 age range and overestimates them over 75 years of age.

### 4.3.6 Age-period term

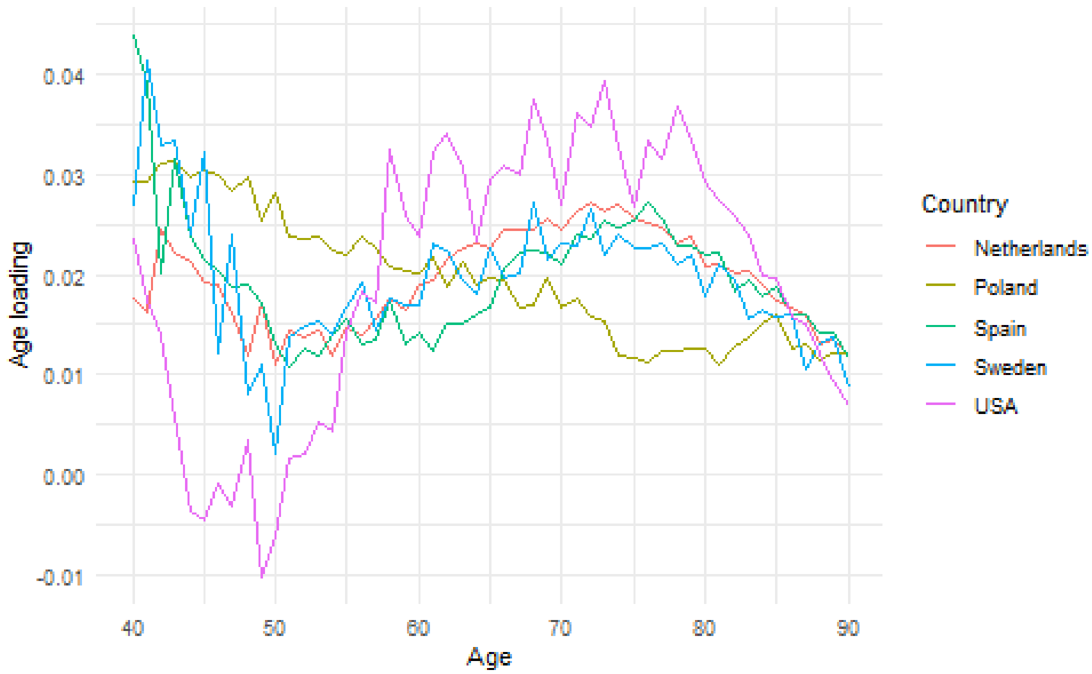


*Figure 4.14: Age loadings $b_x$ of the country-specific residual age-period term $b_x k_t$ in five selected countries*

The age-period term $b_x k_t$ fitted to the data should capture residual trends not explained by the external variables. In an ideal situation, these terms will represent country-specific idiosyncrasies and both the age loadings $b_x$ and the period terms $k_t$ will differ from one country to another. Conversely, similar age loadings and period terms across all countries would point to an omitted variable with a systematic effect on death rates. The age loadings for five selected countries, shown in figure 4.14, present both commonalities and differences. Poland has slightly higher age loadings at younger ages (40-50), which then decline uniformly towards older ages, thus representing middle-aged mortality. The loadings in other countries decrease from age 40 to a trough between ages 45 and 50, with loadings climbing back up, peaking and then declining to age 90. The Netherlands and Spain exhibit a clear peak around age 72 and 75, respectively, while Sweden and the USA show more of a plateau than a simple peak. The shapes of the loadings across countries differs to the point that they eschew a simple, common interpretation. Given the differences in age loadings, the period term $k_t$, shown in figure 4.15, unsurprisingly shows no single common trend, although with some large fluctuations (Poland in 1982) and broken trends (Netherlands 1975-1993-2010).

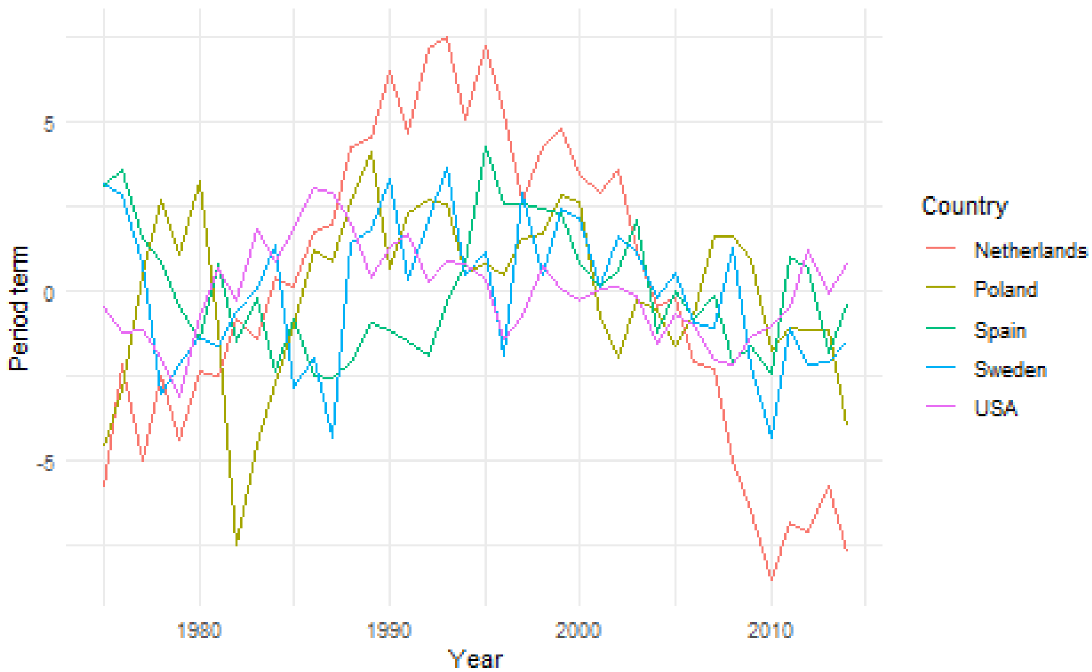


*Figure 4.15: Period term $k_t$ in five selected countries*

### 4.3.7 Composition of external factors

*Table 4.3: Loadings summary for first and second principal component (PC) used in single-population models*

| **Variable** | **1. PC, mean** | **1. PC, st. dev.** | **1. PC, % positive** | **2. PC, mean** | **2. PC, st. dev.** | **2. PC, % positive** |
|---|---|---|---|---|---|---|
| Real GDP | 0.394 | 0.021 | 100.0 | -0.031 | 0.148 | 27.0 |
| Temperature anomalies | 0.241 | 0.096 | 94.6 | -0.014 | 0.286 | 48.6 |
| Fossil fuel consumption | 0.018 | 0.325 | 54.1 | -0.048 | 0.413 | 48.6 |
| Caloric supply | 0.236 | 0.201 | 91.9 | -0.028 | 0.420 | 48.6 |
| Men with raised blood pressure | -0.372 | 0.128 | 2.7 | 0.033 | 0.169 | 54.1 |
| Male height at age 18 | 0.338 | 0.135 | 94.6 | -0.037 | 0.297 | 45.9 |
| Fruit and vegetable consumption | 0.143 | 0.271 | 73.0 | 0.001 | 0.405 | 43.2 |
| Alcohol consumption | -0.129 | 0.308 | 32.4 | -0.011 | 0.352 | 56.8 |
| Cigarette consumption | -0.199 | 0.232 | 21.6 | -0.063 | 0.401 | 43.2 |

While the $O$ external variables $h_{o,t}$ used for the model are the same for all countries, the external factors $g_{l,t}$ actually used in the fitting process are linear combinations of the $h_{o,t}$, obtained through singular value decomposition. The loadings of the principal components express the relative importance of the variables for each principal component and their analysis highlights whether variables' importance is constant or differs across countries.

As shown in table 4.3, the first principal component has a positive loading for real GDP that is quite stable across countries, as can be seen by the low standard deviation of the loading. Likewise, temperature anomalies have a positive loading and a low standard deviation, same for male height at age 18. These three variables all have a positive impact on mortality reductions, while the share of males with raised blood pressure has the opposite impact and has consequently a negative loading.

*Table 4.4: Loadings of the second principal component used in single-population models by country*

| **Country** | **Real GDP** | **Temp. anom.** | **Fossil fuel cons.** | **Caloric supply** | **Men with raised blood press.** | **Male height at age 18** | **Fruit and veget. cons.** | **Alcohol cons.** | **Cigarette cons.** |
|---|---|---|---|---|---|---|---|---|---|
| Australia | -0.042 | 0.014 | 0.122 | **-0.932** | 0.053 | 0.091 | -0.087 | -0.309 | -0.015 |
| Austria | -0.031 | 0.148 | -0.326 | -0.187 | -0.144 | **-0.771** | 0.440 | -0.064 | -0.150 |
| Belgium | 0.223 | -0.486 | 0.179 | **-0.610** | -0.114 | 0.094 | **0.529** | -0.073 | -0.062 |
| Bulgaria | 0.428 | 0.196 | 0.192 | 0.233 | -0.210 | -0.108 | -0.228 | **0.603** | -0.465 |
| Belarus | -0.023 | **0.804** | -0.056 | -0.363 | -0.010 | -0.022 | 0.063 | 0.335 | -0.318 |
| Canada | -0.008 | -0.336 | **-0.862** | -0.092 | -0.092 | 0.120 | 0.019 | -0.307 | -0.129 |
| Chile | -0.189 | 0.246 | 0.184 | 0.106 | 0.025 | 0.063 | 0.008 | **-0.773** | **0.507** |
| Croatia | -0.339 | 0.136 | -0.488 | **-0.545** | 0.447 | -0.082 | -0.088 | 0.233 | -0.251 |
| Switzerland | -0.138 | 0.147 | 0.329 | **-0.813** | 0.048 | 0.037 | -0.389 | -0.154 | 0.109 |
| Czechia | -0.175 | 0.229 | -0.300 | **-0.529** | 0.126 | -0.274 | -0.374 | -0.097 | **-0.554** |
| East Germany | -0.054 | -0.025 | 0.160 | 0.349 | 0.081 | 0.077 | **0.900** | 0.025 | 0.164 |
| West Germany | -0.054 | -0.025 | 0.160 | 0.349 | 0.081 | 0.077 | **0.900** | 0.025 | 0.164 |
| Denmark | -0.018 | -0.344 | 0.405 | 0.374 | -0.049 | 0.335 | -0.107 | 0.440 | **-0.508** |
| Spain | 0.009 | -0.103 | -0.247 | **-0.508** | -0.114 | -0.113 | **-0.595** | 0.033 | **-0.538** |
| Estonia | -0.032 | -0.496 | 0.171 | 0.013 | 0.006 | 0.145 | -0.026 | -0.185 | **-0.817** |
| Finland | -0.010 | 0.322 | **-0.774** | 0.388 | -0.053 | -0.306 | 0.053 | -0.042 | 0.214 |

| Country | Real GDP | Temp. anom. | Fossil fuel cons. | Caloric supply | Men with raised blood press. | Male height at age 18 | Fruit and veget. cons. | Alcohol cons. | Cigarette cons. |
|---|---|---|---|---|---|---|---|---|---|
| France | -0.234 | 0.110 | -0.090 | **0.746** | 0.238 | 0.006 | 0.115 | -0.147 | **0.526** |
| Greece | -0.109 | 0.132 | -0.282 | **-0.566** | -0.106 | -0.151 | **-0.545** | -0.148 | -0.466 |
| Hungary | 0.135 | 0.100 | 0.286 | 0.332 | -0.117 | 0.074 | **0.857** | 0.116 | -0.118 |
| Ireland | -0.051 | -0.398 | -0.426 | -0.191 | -0.142 | -0.012 | 0.255 | **-0.686** | -0.256 |
| Iceland | -0.161 | -0.017 | **-0.771** | 0.305 | 0.160 | -0.499 | -0.088 | 0.062 | 0.027 |
| Italy | -0.025 | -0.216 | **0.560** | 0.335 | 0.305 | -0.043 | **0.520** | 0.067 | 0.397 |
| Japan | 0.025 | -0.127 | -0.041 | **0.716** | 0.120 | 0.420 | 0.126 | 0.301 | 0.414 |
| Republic of Korea | -0.132 | **-0.748** | 0.058 | 0.035 | 0.063 | -0.110 | **-0.508** | 0.359 | -0.125 |
| Lithuania | -0.062 | 0.417 | -0.335 | 0.229 | -0.040 | -0.190 | **0.650** | 0.055 | -0.440 |
| Latvia | 0.171 | 0.100 | -0.087 | -0.170 | -0.037 | 0.275 | -0.336 | 0.034 | **-0.856** |
| Netherlands | 0.123 | 0.105 | **0.919** | 0.116 | 0.061 | -0.141 | -0.068 | 0.127 | 0.269 |
| Norway | -0.107 | -0.207 | 0.337 | -0.069 | 0.042 | **0.508** | -0.114 | -0.416 | **0.617** |
| Poland | 0.227 | -0.085 | 0.155 | **0.544** | -0.171 | -0.133 | -0.103 | **0.677** | -0.326 |
| Portugal | -0.014 | -0.061 | 0.237 | 0.218 | **0.639** | 0.105 | 0.198 | 0.189 | **0.631** |
| Russia | -0.228 | 0.028 | -0.144 | -0.068 | 0.085 | **-0.670** | -0.073 | **0.577** | 0.356 |
| Slovakia | -0.199 | 0.155 | 0.033 | **-0.673** | 0.073 | -0.054 | **-0.638** | -0.234 | -0.115 |
| Slovenia | -0.092 | 0.283 | **-0.609** | -0.093 | -0.038 | 0.029 | -0.389 | 0.130 | **-0.601** |
| Sweden | -0.090 | -0.160 | -0.025 | -0.189 | 0.129 | **0.681** | -0.269 | **-0.609** | 0.075 |
| Ukraine | 0.057 | -0.209 | **0.703** | 0.377 | 0.023 | -0.145 | -0.172 | 0.037 | **0.514** |
| United Kingdom | -0.067 | -0.150 | **-0.566** | -0.168 | -0.069 | -0.038 | 0.001 | **-0.787** | 0.009 |
| USA | 0.150 | -0.007 | **-0.547** | -0.036 | -0.093 | **-0.651** | -0.399 | 0.185 | -0.225 |

The variable loadings for the second principal component are more varied, with most variables more or less equally split between positive and negative loadings. A more detailed look at the loadings in table 4.4 reveals how the second principal component usually has one or two (in the case of Spain, three) variables with a high loading, over 0.5 in absolute value. Cigarette consumption and caloric supply have a loading over 0.5 in absolute value for the second principal component for 11 countries, then fruit and vegetable consumption (10 countries), fossil fuel consumption (9), alcohol consumption (7), male height at age 18 (5), temperature anomalies (2), share of men with raised blood pressure (1), while real GDP per capita never has a loading over 0.43 in absolute value.

## 4.4 Robustness checks and alternative specifications

A number of alternative specifications of the SEV model has been tested, as well as various covariate sets.

While the number of external variables has no direct effect on model complexity, since the external factors are principal components of the SVD of the external variables matrix, it has been tested nevertheless whether excluding certain variables improves model fit by reducing noise. The tests have been carried out for models covering the 40-90 age range, with a maximum of two principal components and an additional age-period term. The tested sets are presented in table 4.5. No set is clearly superior in terms of MAD and MAPE, while considering BIC the full variable set performs best, minimizing the BIC for 11 countries and being the second best choice for 4 more countries. It is therefore possible to further improve on the results presented in section 4.1 by choosing a different subset of covariates, as the full variable set was chosen for all countries for comparability.

*Table 4.5: Alternative variable sets to the base set and number of countries for which the set maximizes MAD and MAPE and minimizes BIC*

| Variable set | Excluded variables | Countries with maximum MAD | Countries with maximum MAPE | Countries with minimum BIC |
|---|---|---|---|---|
| 1 | None | 5 | 6 | 11 |
| 2 | Temperature anomalies | 3 | 2 | 4 |
| 3 | Temperature anomalies and non-stochastic variables (blood pressure, height) | 4 | 4 | 2 |
| 4 | Temperature anomalies and height | 4 | 6 | 5 |
| 5 | Temperature anomalies and blood pressure | 2 | 0 | 0 |
| 6 | Temperature anomalies, blood pressure and fruit/vegetable consumption | 4 | 6 | 1 |
| 7 | Temperature anomalies, blood pressure and fossil fuel consumption | 3 | 4 | 2 |
| 8 | Temperature anomalies and alcohol consumption | 6 | 3 | 8 |
| 9 | Temperature anomalies and cigarette consumption | 6 | 5 | 4 |

Base set of variables: real GDP per capita, temperature anomalies, fossil fuel consumption, caloric

| Variable set | Excluded variables | Countries with maximum MAD | Countries with maximum MAPE | Countries with minimum BIC |
|---|---|---|---|---|
| | supply, share of men with raised blood pressure, average height at age 18, fruit and vegetable consumption per capita, alcohol consumption per capita, cigarette consumption per capita. MAPE not finite for Iceland. | | | |

For both the 0-90 and 40-90 age ranges it has been investigated whether the inclusion of an age-period term and different limitations on the number of external factors can improve model fit. These tests have been carried out with the full covariate set and the results are presented in table 4.6. While more relaxed criteria on inclusion of principal components of external variables increase fit, the BIC suggests more stringent limits. For the 0-90 age range, BIC is minimized for most countries with the inclusion of 3 principal components, while for 6 countries the BIC is minimized with the inclusion of a fourth principal component. For the 40-90 age range, two to three principal components are generally sufficient and additional components add little value. An age-period term markedly improves model fit in all cases. Ultimately, the 40-90 age range has been chosen due to an overall better forecasting performance compared to the Niu-Melenberg model, with the RMFSE ratio being lower than 1 for a higher number of countries at all jump-off years tested.

Model specification has a larger effect on the BIC than the choice of covariates: while testing alternative specifications, the BIC would vary substantially in terms of countries outperforming the corresponding Lee-Carter model, while as far as the choice of variables is concerned, once a model outperformed the Lee-Carter model, it would continue doing so under all sets of variables tested for almost all countries.

*Table 4.6: Parameter sets and number of countries for which the set maximizes MAD and MAPE and minimizes BIC*

| Par. set | Age range | Max n. PC | Variance threshold | Age-period term | Countries with maximum MAD | Countries with maximum MAPE | Countries with minimum BIC |
|---|---|---|---|---|---|---|---|
| A | 0-90 | 5 | 0.10 | Yes | 36 | 25 | 13 |
| B | 0-90 | 3 | 0.10 | Yes | 1 | 0 | 22 |
| C | 0-90 | 5 | 0.10 | No | 0 | 0 | 2 |
| D | 40-90 | 5 | 0.10 | Yes | 37 | 36 | 4 |
| E | 40-90 | 5 | 0.10 | No | 0 | 0 | 1 |
| F | 40-90 | 3 | 0.10 | Yes | 0 | 0 | 6 |
| G | 40-90 | 3 | 0.15 | Yes | 0 | 0 | 11 |
| H | 40-90 | 2 | 0.15 | Yes | 0 | 0 | 15 |

| Par. set | Age range | Max n. PC | Variance threshold | Age-period term | Countries with maximum MAD | Countries with maximum MAPE | Countries with minimum BIC |
|---|---|---|---|---|---|---|---|

MAPE not finite for twelve countries in age range 0-90, for Iceland in age range 40-90. 7 out of 13 countries have 3 principal components above the variance threshold and their BIC is therefore the same as for parameter set B.

### 4.4.1 Cohort terms

*Table 4.7: Mean absolute deviation (MAD), mean absolute percentage error (MAPE), Bayesian Information Criterion (BIC) for single-population models with cohort term and no age-period term*

| Country | MAD | MAPE | BIC |
|---|---|---|---|
| Spain | 81.653 | 1.874 | 27,110 |
| Poland | 100.395 | 3.369 | 33,432 |
| USA | 281.718 | 1.653 | 39,709 |

As shown in the previous sections, some countries exhibit cohort effects, therefore it's sensible to investigate whether including a cohort term would improve the fit. Unfortunately, when fitting the model with both an age-period term and a cohort term, an infinite deviance is produced, preventing a successful estimate of the parameters, hence the inclusion of a cohort term requires the omission of the age-period term.

A single-population model with a cohort term and no age-period term has been estimated for three countries where model residuals suggested a cohort effect: Spain, Poland and the USA, the results are presented in table 4.7. The inclusion of a cohort term yields mixed results for the USA, where there is a decrease in MAPE and BIC, but not in MAD, while for Spain and Poland the fit is noticeably worse.

### 4.4.2 Stationarity of the age-period term

The age-period term $b_x k_t$ fits trends not captured by the external factors $g_{l,t}$. But it's possible that the age-period term captures temporary idiosyncrasies, which vanish in the long term. To this end, the stationarity of the $k_t$ terms has been investigated with the Phillips-Perron test. For 11 countries, the period term $k_t$ is stationary (with a p-value less than 0.05) and therefore for these countries there are no systematic trends not captured by the external factors $g_{l,t}$.

# 5 Discussion

The effect of economic development, environmental and lifestyle factors on mortality is well documented. Prosperity, living in a clean environment and having healthy habits all contribute to a long life. Mortality and these factors are correlated and cointegration analysis has shown that this relationship is stable over the long run even if the individual series are usually non-stationary and also include structural breaks. Hence, external variables can be credibly incorporated in a stochastic mortality model. On one hand, this can improve fit and forecasting performance compared to the Lee-Carter model and to the Niu-Melenberg model which only includes GDP. In addition to improving model performance, including a wider set of variables can also improve the interpretability of stochastic mortality models and offer insights on the relative importance of different factors and how they vary across countries. This makes it easier to elaborate scenarios considering the trade-offs, i.e. between economic growth and environmental protection.

The main goal of this study was to build a model that could be useful in scenario building and policy planning. To this end, expanding the scope of variables included in the model is crucial. If we don't model a variable, it is ignored: the optimal policy to reduce mortality based on a model which only incorporates GDP is to maximize GDP, no matter the effect on environment and on public health, which is clearly nonsensical.

Mortality is a complex phenomenon and stochastic mortality modeling is in a constant trade-off between simple, parsimonious models and more comprehensive models that capture the multifaceted nature of mortality. The complexity of mortality is reflected in the fact that in none of the analysed countries the relationship between mortality and the other variables could be described by a single principal component. To put it otherwise, economic prosperity is not enough. Moreover, while GDP features prominently in the first principal component in all countries, the composition of the second principal component is much more varied, suggesting that different factors have different weight in explaining and predicting country-level mortality. Caloric supply, cigarette consumption, fossil fuel consumption, a proxy for air quality, fruit and vegetable consumption and, to a lesser degree, alcohol consumption feature prominently in the second principal component.

The differences in the weights of the covariates between countries suggest that each country, with its own characteristics and evolution of mortality, faces different challenges. The present analysis, with a fixed set of covariates for all countries, is meant to show the potential benefits of a more flexible model. The SEV model with the full set of external variables outperforms the Lee-Carter model for 18 countries, the Niu-Melenberg model for 12 countries and both models for a diverse set of 11 countries, that is, Australia, Bulgaria, Hungary, Italy, Japan, Netherlands, Russia, Spain, Ukraine, United Kingdom and USA, plus Canada for which a lack of data for a covariate prevents the calculation of the Lee-Carter model for the same years. The choice of the variable set has been shown to have an impact on the BIC and the full variable set minimizes the BIC for just 11 countries (4 of which also outperform the Lee-Carter and Niu-Melenberg models). One can reasonably conclude that for each individual country, both fit and forecasting performance can be improved upon with different sets of covariates, tailored on that country's characteristics, and an appropriate variable selection.

Nevertheless, the choice of covariates appears to have captured the most common factors influencing mortality. Beyond the external variables, the additional age-period term has no clear trend across countries and doesn't have a common pattern for the age loadings either. Its interpretation varies from country to country, suggesting no easily identifiable omitted variables. To the extent where the age loadings peak at a moderately old age (65 years in Italy, 75 in Spain), it could conceivably be interpreted as the effect of the healthcare system's ability to reduce mortality beyond national affluence and lifestyle choices. For a little under a third of all countries considered, though, it's a stationary process and thus it may simply capture past idiosyncrasies with no relevance for forecasting. This suggests that the external factors used actually captured the relevant trends in mortality.

The inclusion of a diverse set of covariates can improve forecasting performance upon the Niu-Melenberg model, as shown in backtesting. Arguably, though, the biggest advantage of using covariates covering a number of different factors is that it allows for wide-ranging model-based scenario planning based on actual historical data. That said, historical correlations are not guaranteed to remain valid in the future as well: the external variables are proxies of actual factors that influence mortality and the relationship between the proxy and the factor may not hold in time, there could be possible non-linear effects and other unforeseen factors that may influence mortality may emerge (i.e. vaping as an alternative to smoking).

In some countries model residuals evidenced the presence of cohort effects, a well-known phenomenon in the study of mortality. The model presented allows for the inclusion of cohort terms at the expense of age-period terms. In general, a cohort term can improve model fit, but it is not a given and its inclusion should be considered carefully against the alternatives and considering the estimation difficulties.

A technical limitation of the model is its need to use uncorrelated factors, which leads to the use of principal components and therefore limits the direct interpretability of the model coefficients. Nevertheless, the impact of a specific variable can be derived from the model coefficients and this doesn't limit the model ability to forecast mortality rates using arbitrary future values of the external variables, which is its main application. Another limitation of stochastic mortality models is their need of data for a relatively large number of years, which limits the set of available covariates and raises questions about the comparability of data across space and time.

A possible avenue for future research would be an extension to the multipopulation case. On the same note, further research is needed to determine why the importance of the different variables in the single-population model varies across countries.

## 6 Acknowledgements

The author would like to thank prof. Pietro Millossovich and prof. Susanna Zaccarin for their helpful comments and the anonymous reviewers for their help in improving the article.

## Bibliography

Ahmad, Omar B, Cynthia Boschi-Pinto, Alan D Lopez, Christopher JL Murray, Rafael Lozano, Mie Inoue, et al. (2001), Age Standardization of Rates: A New WHO Standard, *Geneva: World Health Organization*, 9 (10) (Citeseer): 1–14.

Bai, Jushan and Pierre Perron (2003), Computation and Analysis of Multiple Structural Change Models, *Journal of applied econometrics*, 18 (1) (Wiley Online Library): 1–22.

Berkum, Frank van, Katrien Antonio and Michel Vellekoop (2016), The Impact of Multiple Structural Changes on Mortality Predictions, *Scandinavian Actuarial Journal*, 2016 (7) (Taylor & Francis): 581–603.

Birchenall, Javier A. (2007), Economic Development and the Escape from High Mortality, *World Development*, 35 (4): 543–568.

Boonen, Tim J. and Hong Li (2017), Modeling and Forecasting Mortality with Economic Growth: A Multipopulation Approach, *Demography*, 54 (5): 1921–1946.

Boumezoued, Alexandre, Jean-Baptiste Coulomb, Al Klein, Damien Louvet and Eve Titon (2019), Modeling and Forecasting Cause-of-Death Mortality, *Society of Actuaries Research Report*, 2019 : 1–71.

Bozzo, Giuseppina, Susanna Levantesi and Massimiliano Menzietti (2021), Longevity Risk and Economic Growth in Sub-Populations: Evidence from Italy, *Decisions in Economics and Finance*, 44 (1): 101–115.

Brenner, M Harvey (2005), Commentary: Economic Growth Is the Basis of Mortality Rate Decline in the 20th Century—Experience of the United States 1901–2000, *International Journal of Epidemiology*, 34 (6): 1214–1221.

Cairns, Andrew J. G., David Blake, Kevin Dowd, Guy D. Coughlan, David Epstein and Marwa Khalaf-Allah (2011), Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models, *Insurance: Mathematics and Economics*, 48 (3): 355–367.

Cairns, Andrew J.G., Malene Kallestrup-Lamb, Carsten Rosenskjold, David Blake and Kevin Dowd (2019), Modelling Socio-Economic Differences in Mortality Using a New Affluence Index, *ASTIN Bulletin*, 49 (3) (Cambridge University Press): 555–590.

Danesi, Ivan Luciano, Steven Haberman and Pietro Millossovich (2015), Forecasting Mortality in Subpopulations Using Lee–Carter Type Models: A Comparison, *Insurance: Mathematics and Economics*, 62 : 151–161.

Dutton, Lydia, Athanasios A. Pantelous and Malgorzata Seklecka (2020), The Impact of Economic Growth in Mortality Modelling for Selected OECD Countries, *Journal of Forecasting*, 39 (3): 533–550.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), The Next Generation of the Penn World Table, *American Economic Review*, 105 (10): 3150–82.

French, Declan and Colin O'Hare (2014), Forecasting Death Rates Using Exogenous Determinants, *Journal of Forecasting*, 33 (8): 640–650.

Hanewald, Katja (2011), Explaining Mortality Dynamics, *North American Actuarial Journal*, 15 (2) (Routledge): 290–314.

Holmes, Alexander J. and Kym Anderson (2017), Annual Database of National Beverage Consumption Volumes and Expenditures, 1950 to 2015, (Wine Economics Research Centre, University of Adelaide).
https://universityofadelaide.box.com/s/7xvamsb36f6gpgt9e5l7wjxjvf1uq4g1.

Johansen, Søren (1991), Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica*, 59 (6) ([Wiley, Econometric Society]): 1551–1580.

King, Gary and Samir Soneji (2011), The Future of Death in America., *Demographic Research*, 25 : 1–38.

Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt and Yongcheol Shin (1992), Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?, *Journal of Econometrics*, 54 (1): 159–178.

Lee, Ronald D. and Lawrence R. Carter (1992), Modeling and Forecasting US Mortality, *Journal of the American statistical association*, 87 (419) (Taylor & Francis): 659–671.

Li, Hong and Yanlin Shi (2021), Forecasting Mortality with International Linkages: A Global Vector-Autoregression Approach, *Insurance: Mathematics and Economics*, 100 : 59–75.

(NCD-RisC) (2020), Height and Body-Mass Index Trajectories of School-Aged Children and Adolescents from 1985 to 2019 in 200 Countries and Territories: A Pooled Analysis of 2181 Population-Based Studies with 65 Million Participants., *Lancet (London, England)*, 396 : 1511–1524.

(NCD-RisC), NCD Risk Factor Collaboration (2016), A Century of Trends in Adult Human Height, Franco, Eduardo (Ed.) *eLife*, 5 (eLife Sciences Publications, Ltd): e13410.

Niu, Geng and Bertrand Melenberg (2014), Trends in Mortality Decrease and Economic Growth, *Demography*, 51 (5): 1755–1773.

O'Hare, Colin and Youwei Li (2012), Explaining Young Mortality, *Insurance: Mathematics and Economics*, 50 (1): 12–25.

Perron, Pierre (1989), The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, *Econometrica: journal of the Econometric Society*, (JSTOR): 1361–1401.

Phillips, Peter CB and Pierre Perron (1988), Testing for a Unit Root in Time Series Regression, *Biometrika*, 75 (2) (Oxford University Press): 335–346.

Plat, Richard (2009), On Stochastic Mortality Modeling, *Insurance: Mathematics and Economics*, 45 (3): 393–404.

Poirier, Mathieu JP, G Emmanuel Guindon, Lathika Sritharan and Steven J Hoffman (2019), International Cigarette Consumption Database v1.3, (Borealis).

Preston, Samuel H. (1975), The Changing Relation Between Mortality and Level of Economic Development, *Population Studies*, 29 (2) (Routledge): 231–248.

Riley, James C. (2005), Estimates of Regional and Global Life Expectancy, 1800–2001, *Population and Development Review*, 31 (3): 537–543.

Seklecka, Malgorzata, Norazliani Md. Lazam, Athanasios A. Pantelous and Colin O'Hare (2019), Mortality Effects of Economic Fluctuations in Selected Eurozone Countries, *Journal of Forecasting*, 38 (1): 39–62.

Stoeldraijer, Lenny, Coen van Duin, Leo van Wissen and Fanny Janssen (2013), Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands, *Demographic Research*, 29 (13): 323–354.

Tapia Granados, José A. (2008), Macroeconomic Fluctuations and Mortality in Postwar Japan, *Demography*, 45 (2): 323–343.

Tapia Granados, José A. and Edward L. Ionides (2011), Mortality and Macroeconomic Fluctuations in Contemporary Sweden, *European Journal of Population / Revue européenne de Démographie*, 27 (2): 157–184.

Tapia Granados, José A. and Edward L. Ionides (2017), Population Health and the Economy: Mortality and the Great Recession in Europe, *Health Economics*, 26 (12): e219–e235.

Villegas, Andrés M. and Steven Haberman (2014), On the Modeling and Forecasting of Socioeconomic Mortality Differentials: An Application to Deprivation and Mortality in England, *North American Actuarial Journal*, 18 (1) (Routledge): 168–193.

Villegas, Andrés M., Vladimir K. Kaishev and Pietro Millossovich (2018), StMoMo: An R Package for Stochastic Mortality Modeling, *Journal of Statistical Software*, 84 (3): 1–38.

Wang, Haidong, Mohsen Naghavi, Christine Allen, Ryan M Barber, Zulfiqar A Bhutta, Austin Carter, et al. (2016), Global, Regional, and National Life Expectancy, All-Cause Mortality, and Cause-Specific Mortality for 249 Causes of Death, 1980–2015: A Systematic Analysis for the Global Burden of Disease Study 2015, *The Lancet*, 388 (10053): 1459–1544.

Zeileis, Achim, Ajay Shah and Ila Patnaik (2010), Testing, Monitoring, and Dating Structural Changes in Exchange Rate Regimes, *Computational Statistics & Data Analysis*, 54 (6): 1696–1706.

Zhou, Bin, James Bentham, Mariachiara Di Cesare, Honor Bixby, Goodarz Danaei, Melanie J Cowan, et al. (2017), Worldwide Trends in Blood Pressure from 1975 to 2015: A Pooled Analysis of 1479 Population-Based Measurement Studies with 19.1 Million Participants, *The Lancet*, 389 (10064) (Elsevier): 37–55.