*DEMOGRAPHIC RESEARCH*

*Research Article*

# Coherent forecasts of mortality with compositional data analysis

**Marie-Pier Bergeron-Boucher**

**Vladimir Canudas-Romo**

**Jim Oeppen**

**James W. Vaupel**

# Contents

# Coherent forecasts of mortality with compositional data analysis

**Marie-Pier Bergeron-Boucher**[1]

**Vladimir Canudas-Romo**[2]

**Jim Oeppen**[3]

**James W. Vaupel**[4]

## Abstract

**BACKGROUND**
Mortality trends for subpopulations, e.g., countries in a region or provinces in a country, tend to change similarly over time. However, when forecasting subpopulations independently, the forecast mortality trends often diverge. These divergent trends emerge from an inability of different forecast models to offer population-specific forecasts that are consistent with one another. Nondivergent forecasts between similar populations are often referred to as "coherent."

**METHODS**
We propose a new forecasting method that addresses the coherence problem for subpopulations, based on Compositional Data Analysis (CoDa) of the life table distribution of deaths. We adapt existing coherent and noncoherent forecasting models to CoDa and compare their results.

**RESULTS**
We apply our coherent method to the female mortality of 15 Western European countries and show that our proposed strategy would have improved the forecast accuracy for many of the selected countries. The results also show that the CoDa adaptation of commonly used models allows the rates of mortality improvements (RMIs) to change over time.

**CONTRIBUTION**
This study opens a discussion about the use of age-specific mortality indicators other than death rates to forecast mortality. The results show that the use of life table deaths

[1] Max Planck Odense Center on the Biodemography of Aging, Institute of Public Health, University of Southern Denmark, Odense, Denmark. E-Mail: mpbergeron@health.sdu.dk.
[2] Max Planck Odense Center on the Biodemography of Aging, Institute of Public Health, University of Southern Denmark, Odense, Denmark. E-Mail: vcanudas@health.sdu.dk.
[3] Max Planck Odense Center on the Biodemography of Aging, Institute of Public Health, University of Southern Denmark, Odense, Denmark. E-Mail: joeppen@health.sdu.dk.
[4] Max Planck Institute für Demographic Research, Rostock, Germany. E-Mail: jvaupel@health.sdu.dk.

and CoDa leads to less biased forecasts than more commonly used forecasting models based on the extrapolation of death rates. To the authors' knowledge, the present study is the first attempt to forecast coherently the distribution of deaths of many populations.

# 1 Introduction

Accurate life expectancy forecasts are crucial inputs for decision-making by individuals and by financial, social, and health care institutions. The best way to obtain accurate forecasts is still debated. Different methods for forecasting mortality have been introduced over the years. Booth and Tickle (2008) classify mortality forecasting models into three broad approaches: expert judgment, extrapolation of past trends, and epidemiological models. Most recent developments in forecasting mortality focus on extrapolative models (Booth and Tickle 2008). The extrapolative approach generally finds its robustness in the linear changes over time of different indicators used for forecasts and the limited subjective judgment required (Stoeldraijer et al. 2013; Booth and Tickle 2008; Booth et al. 2006; Oeppen and Vaupel 2002).

A well-known extrapolative approach is the Lee–Carter model (Lee and Carter 1992). The Lee–Carter model uses linear extrapolations of the logarithms of age-specific death rates to forecast mortality, using principal component techniques. While the model works reasonably well (Lee and Miller 2001), one of its flaws is its assumption of a constant rate of age-specific mortality improvement over time. This assumption has been shown to be inadequate in many cases, especially at higher ages, and to overestimate the future level of mortality (Booth, Maindonald, and Smith 2002; Booth and Tickle 2008; Kannisto et al. 1994).

More recently, Oeppen (2008) suggested forecasting life table deaths using Compositional Data Analysis (CoDa), a method pioneered by Aitchison (1986). Compositional data is defined as positive values summing up to a fixed constant, which carry only relative information and represent part of a whole. A compositional vector can be, for example, a vector of proportions or percentages. Oeppen (2008) suggested treating life table distributions of deaths ($d_x$) as compositional data and using a principal component approach, similar to the Lee–Carter model, to forecast mortality within a CoDa framework. Within this framework, the data is constrained to vary between two limits (0 and a given constant), which conditions the relationship between the components, such as ages and causes of deaths, and is manifested in their covariance structure (Pawlowsky-Glahn and Egozcue 2006). This last property can represent an important advantage in a forecasting context (Lee 1998). Furthermore, Oeppen (2008) showed that using CoDa to forecast mortality by age and cause of death does not lead to more pessimistic results than forecasting mortality by age only, as found by previous studies based on death rates

(Wilmoth 1995). It has been shown that when forecasts are based on cause-disaggregated measures, mortality forecasts by components tend to be dominated by an increase or slow decrease of certain subgroups, leading to more pessimistic forecasts (Wilmoth 1995). By using CoDa, this limitation can be avoided (Oeppen 2008).

An important issue related to extrapolative approaches to forecasting is that they often do not consider analogous mortality trends for males and females or for countries in a region or provinces/states in a country. Mortality trends are often projected separately, which tends to increase the divergence between groups in the long run, even when using similar extrapolative procedures (Li and Lee 2005; Wilmoth 1995). Coherent forecasts, i.e., nondivergent forecasts, among industrialized countries are justified, since convergence of mortality levels across industrialized countries has been observed since the middle of the 20th century (White 2002; Wilson 2001, 2011; Li and Lee 2005; Oeppen 2006). This occurred as a general process; populations became integrated via communication, transportation, trade, and technology, without however totally eliminating regional specificities (Li and Lee 2005). Considering this convergence, forecasting mortality by single countries becomes less acceptable and coherent forecasts are often necessary (Li and Lee 2005; Schinzinger, Denuit, and Christiansen 2016; Bohk-Ewald and Rau 2017; Hyndman, Booth, and Yasmeen 2013; Raftery et al. 2013; Cairns et al. 2011; Torri and Vaupel 2012).

Among the solutions offered for the regional coherence problem, Li and Lee (2005) suggest modifying the Lee–Carter method by identifying a factor for central tendency for a group of countries and a factor for individual-country trends. Carter and Lee (1992) and Russolillo, Giordano, and Haberman (2011) suggest using a single time-pattern of mortality change for many populations. Following a suggestion by Oeppen and Vaupel (2002), Torri and Vaupel (2012) forecast the best-practice in life expectancy, which has risen at a steady pace since 1840 (Oeppen and Vaupel 2002), and then forecast the gap between countries' life expectancy and the record level. All these methods suggest using a common trend, reflecting a general mortality process, which influences the country-specific mortality.

Coherent regional mortality forecasts within a CoDa framework have not been explored previously. In this article, we refer to CoDa methodology as a forecast performed within the CoDa framework. The main purpose of this study is to explore the use of an added common factor to the CoDa methodology to obtain an improved coherent forecast method based on the forecast of life table deaths.

There are seven sections in this article. The following section on methods summarizes the Lee and Carter (1992) and Li and Lee (2005) models. It also introduces their respective CoDa adaptation. Section three reports the data used and section four interprets the parameters and shows the explained variance and the fitted models. In the fifth section we present the results of the female forecasts for 15 countries as well as compar-

ing previous models with the new proposal. As final sections, a discussion and conclusion are included.

## 2 Methods

We compare four forecasting models: Lee–Carter and CoDa, both with and without a common factor. We briefly describe them here.

### 2.1 The Lee–Carter (LC) model

The Lee and Carter (1992) model is a principal components approach, based on the log-transformed age-specific death rates ($m_{t,x}$). The model is written as:

$$ln(m_{t,x}) - \alpha_x = \kappa_t \beta_x + \epsilon_{t,x}, \tag{1}$$

where $\alpha_x$ is the age-specific log-mortality average, $\kappa_t$ is the level of mortality in year $t$, $\beta_x$ is an age-pattern of mortality change at age $x$ (also interpreted as the rate of mortality improvement once multiplied by the change in $\kappa_t$ as shown in Appendix F), and $\epsilon_{t,x}$ is the error term. The parameters $\kappa_t$ and $\beta_x$ are the normalized first left and right singular vectors of the singular value decomposition (SVD) of the centered matrix $log(m_{t,x}) - \alpha_x$,

$$\kappa_t = u_t s \sum_{x=0}^{120} v_x \tag{2a}$$

$$\beta_x = \frac{v_x}{\sum\limits_{x=0}^{120} v_x}, \tag{2b}$$

where $u_t$ is the first left-singular vector (years, e.g., from 1960 to 2011), $s$ is the leading singular value, and $v_x$ is the first right-singular vector (ages, e.g. from 0 to 120) of the SVD. Lee and Carter (1992) found that $\kappa_t$ changes linearly and can be forecast using a random walk with drift. Other time series methods could also be used.

## 2.2 The Li–Lee model: LC-coherent

Li and Lee (2005) modified the Lee–Carter model to forecast different populations in a coherent way. Their model uses a common factor, representing an average mortality trend for the whole group of countries. The death rates at time $t$, age $x$ and for population $i$, $m_{t,x,i}$, are modeled as

$$log(m_{t,x,i}) - \alpha_{x,i} - \kappa_t \beta_x = k_{t,i} b_{x,i} + \epsilon_{t,x,i}, \tag{3}$$

where $\alpha_{x,i}$ is the average log-mortality at a given age $x$ for population $i$, and $\kappa_t \beta_x$ is the common factor for all populations. The common factor is obtained by applying the ordinary Lee–Carter method to the average mortality of the group, as in equation (1). The term $k_{t,i} b_{x,i}$ represents the SVD components, as presented in equations (2a) and (2b), of the difference between the centered logged death rates of population $i$ and the rates implied by the common factor (Li and Lee 2005). As stated by the authors, this method is "taking advantage of commonalities in [the populations'] historical experience and age patterns, while acknowledging their individual differences in levels, age patterns, and trends." (Li and Lee 2005: 590)

For the model to work, $k_{t,i}$ should each approach some constant (Li and Lee 2005). "In this way, the fitted model will accommodate some continuation of historical convergent or divergent trends for each country before it locks into a constant relative position in the hierarchy of long-term forecasts of group mortality." (Li and Lee 2005: 578) Li and Lee (2005) suggest forecasting $k_{t,i}$ with a random walk without drift or with an autoregressive model (AR) with intercept. The authors, however, noted that the model can fail if $k_{t,i}$ has a trending long-term mean, which would not guarantee that $k_{t,i}$ will reach a constant.

## 2.3 Forecasting with compositional data analysis (CoDa)

A key difference between the LC method and the Compositional Data Analysis (CoDa) model is that the former forecasts the death rates ($m_{t,x}$) while the latter is based on the life table death distribution ($d_{t,x}$). Both estimators can be derived from the other based on the life table relations (for more information on the life table and its indicators, see Preston, Heuveline, and Guillot (2001)). By using $d_{t,x}$, we model and forecast a redistribution of the density of life table deaths, where deaths at young ages are shifted towards older ages.

By using the $d_{t,x}$, one should be aware of the indicator constraint: The values of $d_{t,x}$ can only vary between 0 and the life table radix, and the sum of the deaths by age for one year $t$ should equal the life table radix. CoDa is a full framework to analyze multivariate data in which the components represent parts of a whole (Aitchison 1986;

Pawlowsky-Glahn and Buccianti 2011). Compositions are vectors of components which are strictly positive, carry only relative information, and always sum to a constant (percentage, per thousand, etc.). According to this definition, the life table deaths can be seen as compositional data (Oeppen 2008). Mert et al. (2016) and Lloyd, Pawlowsky-Glahn, and Egozcue (2012) have shown more generally the utility of CoDa in epidemiology and population studies, and we here suggest a concrete application.

Because the $d_{t,x}$ are constrained to sum to the life table radix, the components are enclosed in a subspace where they can only vary between 0 and the radix value. Such a subspace is referred to as a simplex and does not follow the rules of Euclidean geometry, making the use of standard statistical analysis problematic (Aitchison 1986). Unlike unconstrained multivariate statistical analysis, CoDa offers a framework to deal with such a constraint. CoDa provides a set of tools to deal with compositional problems inside the simplex and to move back and forth from the simplex to the "real space" through log-ratio transformations (Aitchison 1986; Egozcue et al. 2003). These transformations are analogous to the logit transform and its inverse used in logistic regression and the Brass relational mortality model (Brass 1971). In this paper, we use the $clr$ transformation defined as the logarithm of the composition divided by its geometric mean:

$$clr(d_{t,x}) = ln\Big(\frac{d_{t,x}}{g_t}\Big), \tag{4}$$

where $g_t$ is the geometric mean of the age-composition at time $t$. Unlike other more standard transformations (e.g., log transformation), the $clr$ transformation preserves the distance between components from the simplex to the real space (Aitchison 1986; Pawlowsky-Glahn and Egozcue 2006; Pawlowsky-Glahn and Buccianti 2011). More details of this methodology are presented in Appendix A.

Values for components, here the ages, within CoDa are not free to vary independently, an aspect that is manifested in their covariance structure (Aitchison 1986; Pawlowsky-Glahn and Egozcue 2006): If the value of a specific component is decreasing over time, values of at least one other component will have to increase to preserve the constant sum. Modeling and forecasting the $d_{t,x}$ can thus be seen as a lifesaving process as defined by Vaupel and Yashin (1987): Saving a life at a specific age will lead to an extra death at a later age.

Oeppen (2008) proposed forecasting the $d_{t,x}$ using Principal Component Analysis (PCA), similar to Lee and Carter's (1992) suggestion for the $m_{t,x}$, but applied to the death distribution in a CoDa framework:

$$clr(d_{t,x} \ominus \alpha_x) = \kappa_t \beta_x + \epsilon_{t,x}, \tag{5}$$

where $\alpha_x$ is the age-specific geometric mean of the $d_{t,x}$ over time; $\kappa_t$ is the time index and $\beta_x$ is the age pattern found by SVD; and $\epsilon_{t,x}$ are the errors. The operator $\ominus$ is a standard CoDa operator and is defined as a perturbation procedure (see details in Appendix A). This operator is used to center the matrix while retaining the constant sum. In the CoDa methodology and based on a rank-1 approximation, the parameters $\kappa_t$ and $\beta_x$ are estimated by

$$\kappa_t = u_t s \tag{6a}$$

$$\beta_x = v_x, \tag{6b}$$

where $u_t$ is the first left-singular vector (years), $s$ is the leading singular value, and $v_x$ is the first right-singular vector (ages) of the SVD. The way to estimate $\kappa_t$ and $\beta_x$ in CoDa differs from that suggested by Lee and Carter (1992). With the model introduced in equation (5), the $clr$ coordinates are double centered (over time and age). This last property makes the normalization suggested by Lee and Carter (1992), shown in equations (2a) and (2b), of reaching a unique solution unnecessary, as $\kappa_t$ and $\beta_x$ are automatically normalized to sum to 0. However, $\kappa_t$ and $\beta_x$ are not unique as their estimates can be symmetric around 0, i.e., sometimes both sets of parameters increase over time or age and sometimes they both decrease. The former case (increase) was found for all countries included in the Results section and is thus considered the standard. If the latter case occurs, the parameters could be multiplied/divided by –1.

Once the parameters are estimated from equation (5), the estimated $d_{t,x}$ are found by

$$d_{t,x} = \alpha_x \oplus C[e^{\kappa_t \beta_x + \epsilon_{t,x}}], \tag{7}$$

where $C[]$ is a closing procedure used to transform the estimates into compositional data summing up to the initial constant. This is equivalent to calculating the proportions in each year $t$. To re-enter compositional data form, following a $clr$ transformation, the inverse $clr$ is used (see Appendix A). This procedure comprises exponentiating the $clr$ coordinates and then closing the result. The $\oplus$ is also a perturbation operator and is here used to reverse the centering perturbation shown in equation (5). The step-by-step approach of equations (5) and (7) is presented in Appendix A. As for the LC model, the time index is forecast using time series methods.

## 2.4 The CoDa-coherent model

In order to reach coherence in mortality forecasts, we suggest modifying the above CoDa model by adding a common factor as suggested by Li and Lee (2005). The common factor is found in equation (7) when applying the CoDa methodology to the average $d_{t,x}$

for a group of populations and is denoted as $C[e^{\kappa_t \beta_x}]$. The CoDa-coherent model can then be written analogously to the Li and Lee (2005) formulations in equation (3):

$$clr(d_{t,x,i} \ominus \alpha_{x,i} \ominus C[e^{\kappa_t \beta_x}]) = k_{t,i} b_{x,i} + \epsilon_{t,x,i}, \tag{8a}$$

or

$$d_{t,x,i} = \alpha_{x,i} \oplus C[e^{\kappa_t \beta_x}] \oplus C[e^{k_{t,i} b_{x,i} + \epsilon_{t,x,i}}]. \tag{8b}$$

As with the LC-coherent model, $\kappa_t \beta_x$ is the common factor for all populations found by applying the CoDa methodology presented by equation (5) to the average mortality of a group of populations. The term $k_{t,i} b_{x,i}$ is the country-specific perturbation factor from the common factor and represents the SVD components presented in equations (6a) and (6b) of the matrix $clr(d_{t,x,i} \ominus \alpha_{x,i} \ominus C[e^{\kappa_t \beta_x}])$.

To avoid diverging trends, $k_{t,i}$ should, as for the LC-coherent model, approach a constant. Different time series models fulfill this criterion. However, as for the LC-coherent model, the CoDa-coherent model cannot guarantee that $k_{t,i}$ will reach a constant, especially if the index is recording a long-term increasing or decreasing trend. In this context, the coherence with other countries might not occur.

The prediction intervals (PI) for all four models presented in sections 2.1 to 2.4, referred to as the LC, LC-coherent, CoDa, and CoDa-coherent models, respectively, are built with a bootstrap process and based on the Keilman and Pham (2006) procedure. The method is detailed in Appendix B.

# 3 Data

The data used in this study comes from the Human Mortality Database (HMD 2016). The HMD offers historical data on mortality. The data series is constructed according to a common protocol, making the HMD an excellent comparative source (Barbieri et al. 2015). The study will focus on forecasting the female mortality of 15 Western European countries: Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

The number of years with available mortality data differs for each country within the HMD. However, the HMD covers the period 1960–2011 for all the selected countries, after combining East and West Germany (see Appendix C). These years will be the reference period to forecast mortality.

We extracted the observed death counts and exposure-to-risk estimates from the HMD and calculated life tables. The calculation of the average mortality of the 15 countries and how we deal with 0 values and data at old ages are explained in Appendix C.

# 4  The underlying models

## 4.1  The parameters' interpretation and forecasts

In the methods section, we used similar notation for the parameters of the LC and CoDa models and their coherent versions, as they have similar interpretations but are not identical. The parameters $\beta_x$, shown in Figure 1, provide an age pattern of the mortality changes. In the LC model, the $\beta_x$ produce the different rates of mortality improvement by age, when multiplying by the time factor $\kappa_t$. In a CoDa model, this parameter indicates the transfer of $d_{t,x}$ from one age to another. The density of deaths for ages where $\beta_x$ are negative will be transferred towards ages where $\beta_x$ are positive in relative terms (Oeppen 2008).

**Figure 1:**     **Age pattern ($\beta_x$) of the Lee–Carter and CoDa models for the average (in black) and country-specific (in grey) female mortality of 15 European countries, 1960–2011**
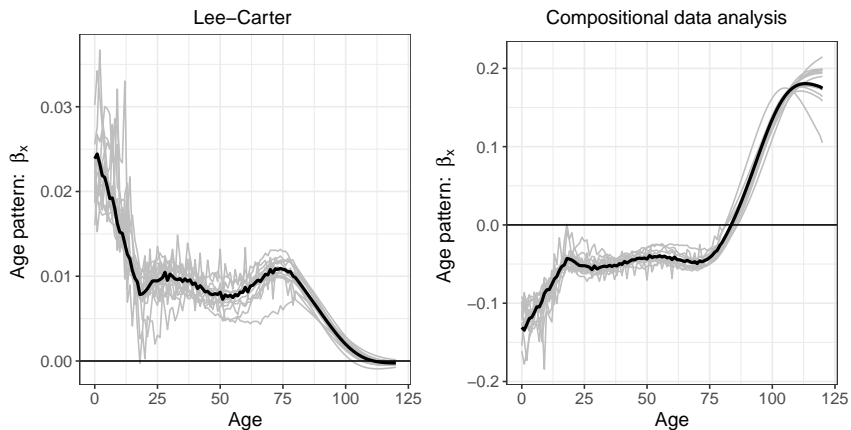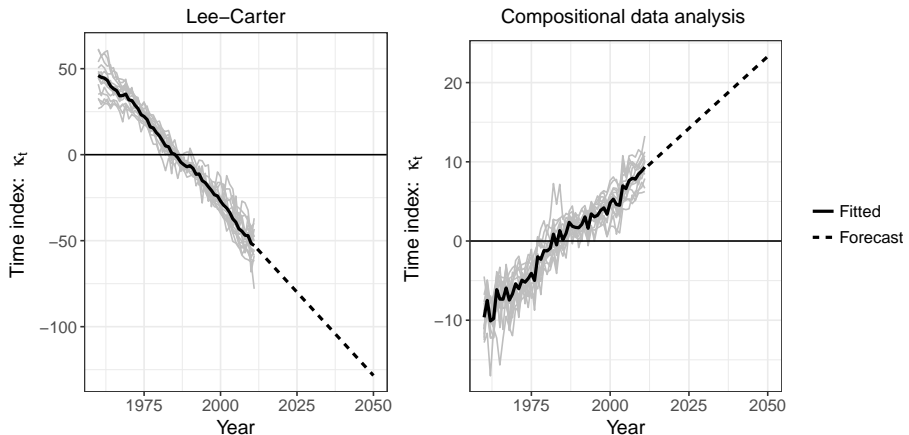


*Source*: HMD (2016) and authors' calculations.

The parameters $\kappa_t$ are indices of the general level of mortality over time for both models. Figure 2 shows that both $\kappa_t$ estimates change linearly over time, although the $\kappa_t$

for the CoDa model have more pronounced fluctuations. The coefficient of determination ($R^2$) of a linear regression applied to the $\kappa_t$ of the average for each of the models is 99.6% for the LC model and of 97.1% for the CoDa model.

**Figure 2:** **Time index ($\kappa_t$) of the Lee–Carter and CoDa models for the average (in black) and country-specific (in grey) female mortality of 15 European countries, observed 1960–2011 and forecast 2012–2050**



*Source*: HMD (2016) and authors' calculations.

In the CoDa model, the ages with negative $\beta_x$ recorded a decrease of their density of deaths over time, in relative terms, once multiplied by $\kappa_t$. They start with a $d_{t,x}$ value higher than the estimated average $\alpha_x$ and this value decreases over time. The $d_{t,x}$ become smaller than $\alpha_x$ when $\kappa_t$ crosses zero. For the ages where the $\beta_x$ are positive, the $d_{t,x}$ is initially lower than the average and then increases over time. In Figures 1 and 2, the density of deaths is thus transferred from younger ages toward older ages.

Lee and Carter (1992) suggest forecasting $\kappa_t$ using a random walk with drift (ARIMA (0,1,0)). We use this procedure to forecast the LC $\kappa_t$. Based on the best BIC value, the $\kappa_t$ of the CoDa model is forecast with an ARIMA(0,1,1) model. This model was the one with the best BIC values for most of the selected countries. We here use the same time series for all countries to introduce and compare our model with existing models. However, other time series models could easily be used when forecasting country-specific mortality with CoDa.

In the present article, we compare four models: LC, LC-coherent, CoDa and CoDa-coherent. For the models considering the coherence between countries (CoDa-coherent and LC-coherent), the time index of the difference between the average and the country-

specific trends, $k_{t,i}$, should also be forecast. Based on the best BIC values for ARIMA models excluding the use of a drift and moving average (MA) components – i.e., selection between random walk without drift and autoregressive model (AR) as suggested by Li and Lee (2005) – an ARIMA(1,1,0) without drift was selected. This model allows the $k_{t,i}$ to reach a constant while fitting the trends of many countries. This procedure was applied to both LC-coherent and CoDa-coherent models.

## 4.2 Explained variance and fitted models

Within a singular value decomposition (SVD), the combination of the first left and right singular vectors and first singular value, called a rank-1 or one-dimensional approximation, is the one explaining the greatest variance. If the explained variance for a rank-1 approximation is low, higher rank approximations can be used. Table 1 presents the explained variance for a rank-1 and rank-2 approximation for the four models and the 15 countries. The Table shows that a rank-1 approximation of the centered matrix of the $m_{t,x}$ with the LC model and of the $d_{t,x}$ with the CoDa model explains most of the variance for most countries. With the CoDa model, a rank-1 approximation explains more than 80% of the variance for 13 out of 15 countries. A rank-2 approximation would increase the explained variance by 7% or less for all countries. Similar results are found for the LC model for most countries. In most cases, no major gains in terms of explained variance would come from adding additional parameters for the second rank in the models. For the coherent models (LC-coherent and CoDa-coherent), the explained variance is lower as it is estimated from the $m_{t,x}$ and $d_{t,x}$ matrices after the common trend has been removed.

The variance explained by the LC model is lower than for the CoDa model for most countries. The errors in modeling and forecasting mortality with the LC and LC-coherent models could then be more important than with the CoDa and CoDa-coherent models.

In Appendix D, we also show that the fit is generally good for most ages and for both CoDa and CoDa-coherent models. The fit was especially good at higher ages, representing an advantage of the CoDa method. Changes in mortality at higher ages have been more influential on life expectancy after 1960 (Bergeron-Boucher, Ebeling, and Canudas-Romo 2015).

**Table 1:** **Explained variance of a rank-1 and rank-2 approximation of a singular value decomposition applied within four models, 15 countries and their average, 1960–2011**

| Country | LC ($\kappa_t\beta_x$) rank-1 | rank-2 | 2-1 | LC-coherent ($k_{t,i}b_{x,i}$) rank-1 | rank-2 | 2-1 | CoDa ($\kappa_t\beta_x$) rank-1 | rank-2 | 2-1 | CoDa-coherent ($k_{t,i}b_{x,i}$) rank-1 | rank-2 | 2-1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Germany | 0.96 | 0.97 | 0.01 | 0.43 | 0.55 | 0.12 | 0.98 | 0.99 | 0.01 | 0.81 | 0.87 | 0.06 |
| Italy | 0.95 | 0.96 | 0.01 | 0.43 | 0.57 | 0.14 | 0.98 | 0.99 | 0.01 | 0.80 | 0.85 | 0.05 |
| France | 0.94 | 0.96 | 0.02 | 0.32 | 0.52 | 0.19 | 0.98 | 0.98 | 0.01 | 0.69 | 0.79 | 0.10 |
| Spain | 0.93 | 0.95 | 0.02 | 0.40 | 0.56 | 0.15 | 0.91 | 0.96 | 0.05 | 0.74 | 0.80 | 0.06 |
| United Kingdom | 0.92 | 0.94 | 0.02 | 0.43 | 0.54 | 0.11 | 0.96 | 0.98 | 0.02 | 0.57 | 0.76 | 0.20 |
| Portugal | 0.88 | 0.91 | 0.03 | 0.38 | 0.52 | 0.14 | 0.93 | 0.95 | 0.03 | 0.57 | 0.71 | 0.14 |
| The Netherlands | 0.82 | 0.84 | 0.03 | 0.32 | 0.43 | 0.11 | 0.86 | 0.93 | 0.06 | 0.75 | 0.80 | 0.05 |
| Belgium | 0.80 | 0.82 | 0.02 | 0.10 | 0.19 | 0.09 | 0.93 | 0.95 | 0.02 | 0.47 | 0.54 | 0.08 |
| Austria | 0.79 | 0.82 | 0.03 | 0.14 | 0.25 | 0.12 | 0.90 | 0.92 | 0.02 | 0.44 | 0.52 | 0.08 |
| Switzerland | 0.69 | 0.73 | 0.04 | 0.13 | 0.23 | 0.10 | 0.86 | 0.88 | 0.02 | 0.28 | 0.36 | 0.09 |
| Finland | 0.68 | 0.71 | 0.03 | 0.12 | 0.21 | 0.09 | 0.88 | 0.91 | 0.03 | 0.63 | 0.68 | 0.05 |
| Sweden | 0.67 | 0.73 | 0.05 | 0.16 | 0.29 | 0.13 | 0.81 | 0.84 | 0.04 | 0.44 | 0.53 | 0.09 |
| Ireland | 0.60 | 0.65 | 0.06 | 0.15 | 0.25 | 0.10 | 0.73 | 0.77 | 0.05 | 0.35 | 0.45 | 0.10 |
| Denmark | 0.55 | 0.61 | 0.06 | 0.16 | 0.27 | 0.11 | 0.80 | 0.83 | 0.03 | 0.28 | 0.39 | 0.11 |
| Norway | 0.48 | 0.57 | 0.08 | 0.16 | 0.28 | 0.13 | 0.66 | 0.73 | 0.07 | 0.36 | 0.46 | 0.10 |
| Average | 0.97 | 0.98 | 0.01 | - | - | - | 0.98 | 0.99 | 0.01 | - | - | - |

*Source*: HMD (2016) and authors' calculations.
*Note*: The countries are listed by order of explained variance obtained with the LC model.

# 5 Results

## 5.1 Evaluating the models

Table 2 shows how well each model could have predicted the mortality age-pattern for the period 1995–2011, based on the reference period 1960–1994. The table presents the mean absolute error (MAE) for the logged age-specific death rates over all ages and years for the four models. The table also shows the Aitchinson distance (AD) of the forecast $d_{t,x}$ in comparison with the observed $d_{t,x}$. The AD is a measure of dissimilarity in CoDa, defined as the square root of the sum of squared difference between two compositions expressed in $clr$ coordinates (more details are provided in Appendix A) (Aitchison et al. 2000; Pawlowsky-Glahn and Buccianti 2011). The coherent versions of the LC and CoDa models would have been more accurate in predicting the age pattern of mortality over the period 1995–2011 than their noncoherent versions. The CoDa-coherent model has the

lowest MAE average across countries for both $m_{t,x}$ and $d_{t,x}$ (0.158 and 2.584 respectively). This last model would have also performed better than the other models in six and seven countries for $m_{t,x}$ and $d_{t,x}$, respectively. However, the LC-coherent model performs better for seven countries for the $m_{t,x}$, with an average of 0.164.

**Table 2:** **Mean absolute error (MAE) of female logged age-specific death rates ($m_{t,x}$) over ages and years and the mean Aitchinson distance (AD) over time for each forecast composition of life table deaths ($d_{t,x}$)**

| Country | MAE: $m_{t,x}$ | | | | AD:$d_{t,x}$ | | | |
|---|---|---|---|---|---|---|---|---|
| | LC | LC-coherent | CoDa | CoDa-coherent | LC | LC-coherent | CoDa | CoDa-coherent |
| United Kingdom | 0.09 | 0.08 | 0.10 | 0.09 | 1.27 | 1.22 | 1.25 | 1.24 |
| The Netherlands | 0.12 | 0.12 | 0.12 | 0.13 | 1.92 | 1.94 | 1.90 | 1.97 |
| France | 0.13 | 0.11 | 0.14 | 0.11 | 1.69 | 1.55 | 1.70 | 1.53 |
| Germany | 0.13 | 0.14 | 0.11 | 0.12 | 1.91 | 1.95 | 1.84 | 1.88 |
| Italy | 0.14 | 0.16 | 0.13 | 0.13 | 2.09 | 2.20 | 2.04 | 2.09 |
| Spain | 0.14 | 0.15 | 0.14 | 0.14 | 2.15 | 2.19 | 2.16 | 2.13 |
| Belgium | 0.14 | 0.13 | 0.15 | 0.14 | 2.28 | 2.23 | 2.26 | 2.23 |
| Portugal | 0.15 | 0.16 | 0.14 | 0.14 | 2.38 | 2.41 | 2.39 | 2.33 |
| Austria | 0.19 | 0.19 | 0.17 | 0.17 | 2.79 | 2.78 | 2.71 | 2.71 |
| Sweden | 0.19 | 0.18 | 0.19 | 0.19 | 2.97 | 2.95 | 2.93 | 2.88 |
| Finland | 0.20 | 0.19 | 0.20 | 0.20 | 3.21 | 3.18 | 3.20 | 3.17 |
| Norway | 0.21 | 0.20 | 0.21 | 0.21 | 4.12 | 4.04 | 4.13 | 4.12 |
| Switzerland | 0.23 | 0.22 | 0.23 | 0.19 | 3.53 | 3.44 | 3.48 | 3.20 |
| Denmark | 0.25 | 0.20 | 0.24 | 0.20 | 3.68 | 3.46 | 3.63 | 3.44 |
| Ireland | 0.25 | 0.23 | 0.25 | 0.22 | 4.00 | 3.82 | 3.91 | 3.84 |
| Mean | 0.17 | 0.16 | 0.17 | 0.16 | 2.67 | 2.62 | 2.63 | 2.58 |
| No. countries | 0 | 7 | 2 | 6 | 0 | 4 | 4 | 7 |

*Source*: HMD (2016) and authors' calculations.
*Note*: This table shows the mean absolute error (MAE) of female logged age-specific death rates ($m_{t,x}$) over ages and years and the mean Aitchinson distance (AD) over time for each forecast composition of life table deaths for the forecast period 1995–2011, based on the reference period 1960–1994, their average over countries, and number of countries recording the lowest MAE and AD by model. The countries are listed by order of MAE for the $m_{t,x}$ obtained with the LC model.

Table 3 presents MAE and the mean error (ME) of female life expectancy at birth forecast for the period 1995–2011, based on the reference period 1960–1994, for 15 countries (see also Appendix E). The MAE is a measure of forecast accuracy while the ME is a measure of bias of the methods. The ME is here defined as: $mean[e_0^{Expected} - e_0^{Observed}]$. The table shows that adding a factor of central tendency in the CoDa forecast model

would have, in general, increased the accuracy of the forecasts. The coherent version of the CoDa model would have performed better than the other models in 5 out of 15 countries, followed by the CoDa model in 4 countries.

**Table 3:      Mean absolute error (MAE) and mean error (ME) of female life expectancy at birth for the forecast period 1995–2011**

| Country | MAE | | | | ME | | | |
|---|---|---|---|---|---|---|---|---|
| | LC | LC-coherent | CoDa | CoDa-coherent | LC | LC-coherent | CoDa | CoDa-coherent |
| Switzerland | 0.11 | 0.17 | 0.80 | 0.42 | 0.09 | 0.08 | 0.80 | 0.42 |
| Belgium | 0.13 | 0.10 | 0.45 | 0.42 | 0.11 | 0.05 | 0.44 | 0.41 |
| Spain | 0.14 | 0.34 | 0.16 | 0.16 | 0.00 | −0.32 | 0.16 | 0.15 |
| Finland | 0.23 | 0.35 | 0.12 | 0.20 | −0.15 | −0.30 | 0.05 | −0.17 |
| France | 0.25 | 0.25 | 0.83 | 0.54 | 0.24 | −0.06 | 0.83 | 0.54 |
| The Netherlands | 0.32 | 0.43 | 0.31 | 0.76 | 0.09 | 0.43 | 0.04 | 0.76 |
| Sweden | 0.40 | 0.32 | 0.64 | 0.69 | 0.40 | 0.32 | 0.64 | 0.69 |
| United Kingdom | 0.42 | 0.22 | 0.25 | 0.20 | −0.32 | −0.02 | −0.08 | 0.19 |
| Italy | 0.44 | 0.57 | 0.22 | 0.20 | −0.44 | −0.57 | 0.18 | 0.13 |
| Germany | 0.48 | 0.52 | 0.17 | 0.19 | −0.48 | −0.52 | −0.11 | −0.17 |
| Norway | 0.54 | 0.17 | 0.34 | 0.44 | −0.54 | −0.08 | −0.30 | 0.44 |
| Portugal | 0.55 | 0.54 | 0.43 | 0.36 | −0.35 | −0.36 | −0.18 | −0.21 |
| Austria | 0.70 | 0.69 | 0.45 | 0.45 | −0.70 | 0.69 | −0.45 | −0.45 |
| Ireland | 0.79 | 0.71 | 0.87 | 0.54 | −0.51 | −0.32 | −0.63 | −0.06 |
| Denmark | 1.06 | 0.27 | 0.60 | 0.20 | −1.02 | 0.01 | −0.54 | 0.17 |
| Mean | 0.44 | 0.38 | 0.44 | 0.38 | −0.24 | −0.16 | 0.06 | 0.19 |
| No. countries | 3 | 3 | 4 | 5 | 9:6 | 10:5 | 7:8 | 5:10 |

*Source*: HMD (2016) and authors' calculations.
*Note*: The table shows the mean absolute error (MAE) and mean error (ME) of female life expectancy at birth for the forecast period 1995–2011, based on the reference period 1960–1994, for 15 countries, their average and number of countries recording the lowest MAE by model or the number of countries with negative vs positive (negative:positive) ME within each model. The countries are listed by order of MAE obtained with the LC model.

In terms of bias, the number of countries with negative vs positive (negative:positive) ME is 9:6 with the LC model, 10:5 with the LC-coherent, 7:8 with the CoDa model, and 5:10 with the CoDa-coherent model. The least biased model is then CoDa. The LC and LC-coherent models tend to underestimate life expectancy, a well-known aspect of the LC model and its variants (Booth, Maindonald, and Smith 2002; Booth and Tickle 2008; Kannisto et al. 1994). On the other hand, the CoDa-coherent model does not seem to preserve the "least-bias" advantage of the CoDa model, and tends to overestimate life expectancy, especially the mortality of the Netherlands. This country had a life expectancy

above average before the jump-off year (1994). The CoDa-coherent model predicted that the Netherlands' life expectancy would stay above the average, but it fell behind, an aspect that the model was not able to anticipate. On the other hand, the model was quite accurate in forecasting the catch-up of Danish females, which started around the jump-off year.

When comparing the accuracy of the two coherent models, LC-coherent and CoDa-coherent, the latter performs better. Using a CoDa-coherent model would have increased the accuracy of the forecast life expectancy for 9 out of 15 countries in comparison with the LC-coherent model (Table 3). However, on average both models have an equal MAE. When looking at the forecast age pattern (Table 2), the CoDa-coherent model would have outperformed the LC-coherent model for 8 and 10 out of 15 countries, for the $m_{t,x}$ and $d_{t,x}$, respectively. As mentioned previously, the CoDa-coherent model has a lower MAE average when estimating the accuracy of the predicted age patterns, both with $m_{t,x}$ and $d_{t,x}$.

Figure 3 shows the life expectancy forecast with the LC-coherent and CoDa-coherent model in 2011, compared with the observed value. The LC-coherent model underestimated the life expectancy of 13 out of 15 countries by the end of the forecast period. The CoDa-coherent model under-predicted the life expectancy of 6 out of 15 countries in 2011 and predicted life expectancy better for 9 countries, in comparison with the LC-coherent model. The prediction intervals (PI) with the CoDa-coherent model are generally wider than with the LC-coherent model. The LC models are generally known to produce very narrow PI (Keilman and Pham 2006). In 2011, the PI of the LC-coherent model contains the actual life expectancy at birth for 86.7% of the countries (13/15) for the 95% coverage; and 66.7% (10/15) of the countries for the 80% coverage. However, the CoDa-coherent model might produce PI that are too wide. In 2011, the CoDa-coherent model contained the actual life expectancy for 100% (15/15) and 86.7% (13/15) of the countries for the 95% and 80% coverage, respectively.

The CoDa-coherent model produced somewhat more accurate life expectancy and age pattern forecasts for the years 1995 to 2011. However, the CoDa-coherent model tended to overpredict life expectancy in this period, especially the life expectancy of the Netherlands. The noncoherent CoDa model is, however, generally less biased.

## 5.2 Life expectancy in 2050

### 5.2.1 More optimistic forecasts

Figure 4 shows female life expectancy at birth forecast for all 15 selected countries with the LC, LC-coherent, CoDa, and CoDa-coherent models. A CoDa approach, with or without a common factor, gives more optimistic forecasts. These more optimistic forecasts come from the fact that the rates of mortality improvement (RMIs) – i.e., the relative

**Figure 3:** **Female life expectancy at birth in 2011 for 15 countries observed and forecast with the LC-coherent and CoDa-coherent models, using 1960–1994 as reference period, and their 80% and 95% PI**
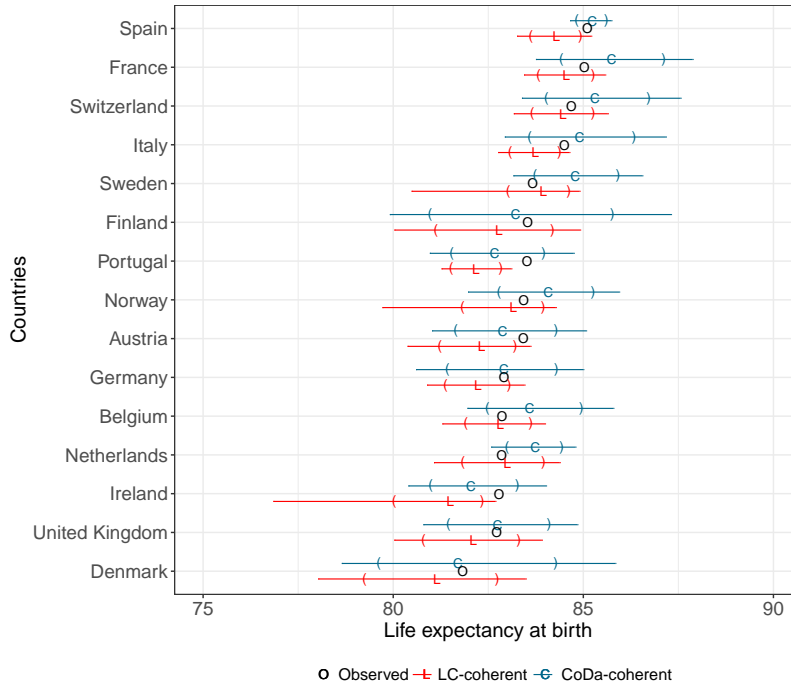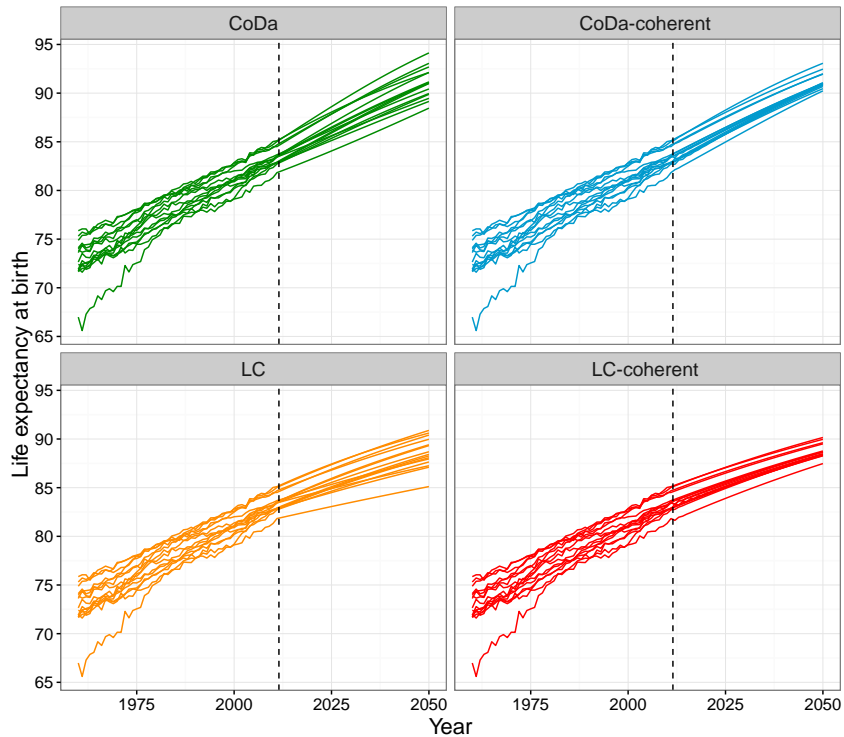


*Source*: HMD (2016) and authors' calculations.
*Notes*: The countries are listed by order of life expectancy observed in 2011.

change in the age-specific death rates from one year to another (see Appendix F for more details on the RMIs) – for the CoDa and CoDa-coherent models can change over time, while they stay constant with the LC and LC-coherent models when $\kappa_t$ is forecast with a random walk with drift, as shown in Appendix F. By using a CoDa methodology, the main LC model problem – i.e., the fixed RMIs (Booth, Maindonald, and Smith 2002; Booth and Tickle 2008; Kannisto et al. 1994) – can then be overcome. As shown in the previous section, the CoDa model is generally less biased. However, the CoDa-coherent model tends to overestimate life expectancy at birth.

Figure 5 shows the life expectancy forecast in 2050 for all 15 countries with the LC-coherent and the CoDa-coherent models with their prediction intervals (PI). The CoDa-coherent model leads to more optimistic forecasts for all countries. As mentioned previ-

**Figure 4:** **Female life expectancy at birth, observed 1960–2011 and forecast 2012–2050 for 15 European countries using four forecasting models**

ously, the PI of the CoDa and CoDa-coherent models are generally wider than with the LC and LC-coherent models, meaning that the uncertainty is greater when forecasting with a CoDa methodology than with an LC model, coherent or not. In 2050, the width of the 95% PI for France is 5.3 years with the CoDa-coherent model and 3.8 with the LC-coherent model. However, as shown in Figure 3, the PI for the LC-coherent model are sometimes very narrow and sometimes do not include the observed values. The wider PI from the CoDa models might come from the more pronounced fluctuations of the CoDa time index as shown in Figure 2, suggesting that the relative residuals after fitting the selected time series model might be more important. Due to the bootstrap process used to calculate the PI (see Appendix B), if the errors, or their extreme values, are more positive

than negative, or vice versa, the PI bounds might be asymmetric. This is, for example, the case of Norway when forecast with the LC-coherent model.

**Figure 5:** **Female life expectancy at birth for 15 countries, observed in 2011 and forecast in 2050 with the LC-coherent and CoDa-coherent models, using 1960–2011 as reference period, and their 80% and 95% PI**
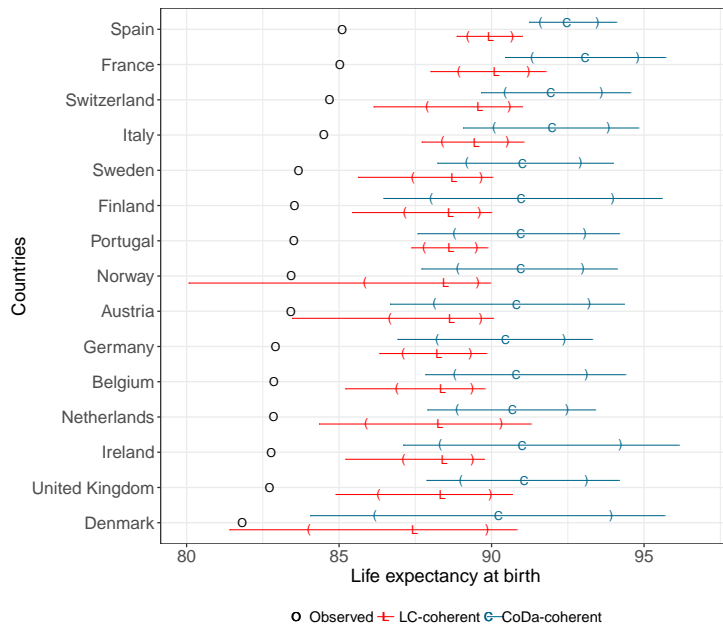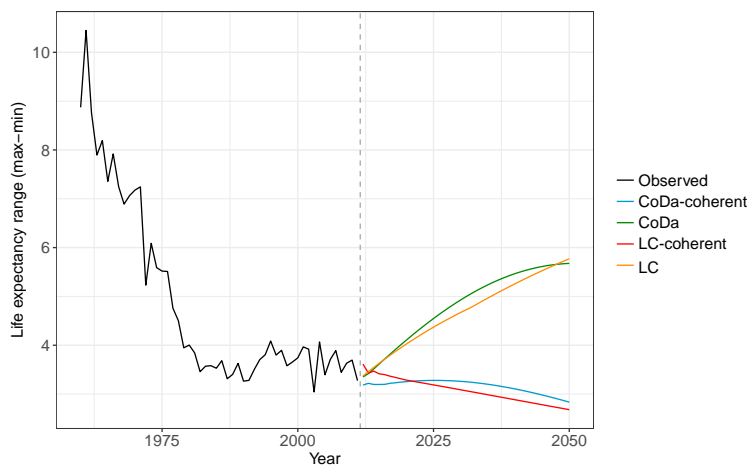


*Source*: HMD (2016) and authors' calculations.
*Notes*: The countries are listed by order of life expectancy observed in 2011.

### 5.2.2 Coherence in the forecasts

Figure 6 shows the range (maximum–minimum values) of life expectancy among the selected 15 countries. Between 1960 and 2011, the range of life expectancy values decrease from 8.87 years to 3.28 years, mainly due to Portugal catching up with the other countries. Since the 1980s, the range of life expectancy values remains around 3.6 years, confirming the need for coherent forecasting among Western European countries.

Figures 4 and 6 show that adding a common factor to the LC and CoDa models succeeds in reducing the long-term divergence in the forecast life expectancy. For example, under the LC model the difference between the maximum and minimum in the forecast life expectancies in 2050 is 5.77 years, while for the LC-coherent model that gap is 2.68 years. Similar results are found when comparing the CoDa and CoDa-coherent models, with ranges of 5.68 and 2.84 respectively. Using a trend common to Western European countries thus allows one to forecast life expectancy in a more coherent way and avoids increasing divergence in the long term. However, the coherent models predict a further convergence, albeit modest, of life expectancy values even if the range stayed approximately constant in the last three decades.

**Figure 6:** **Range of female life expectancy at birth for 15 European countries, observed from 1960–2011 and forecast from 2012–2050 using four methods**



*Source*: HMD (2016) and authors' calculations.

## 6 Discussion

The CoDa methodology is a new forecasting approach and this article is the first to explore its potential to forecast life expectancy coherently among many countries. The results show that using a CoDa-coherent model is a compelling strategy to forecast mortality. One important advantage of the model is the changing RMIs over time, which overcomes the problematic fixed RMI assumption of the LC model. This last aspect of the LC model has been criticized for yielding too pessimistic forecasts (Booth, Maindonald, and Smith 2002; Kannisto et al. 1994). The noncoherent CoDa methodology allows for more optimistic and less biased forecasts for Western European countries. However, the CoDa-coherent model might sometimes be too optimistic in its forecast.

The changes in the RMIs can come from two aspects of the model: 1) the use of the $clr$ transformation, which does not produce constant RMI due to the closing procedure $C[]$ (see Appendix F); and 2) the use of the $d_{t,x}$ as indicator. Due to the relation between indicators in the life table, modeling an indicator in a certain way might lead to different modeling of other life table indicators, producing different RMIs. More detailed analyses should be performed on the consequences of using different indicators for the forecast results.

Despite somewhat more accurate forecasts, the PI are wider with a CoDa method – suggesting that the forecasts carry more uncertainty – than with an LC method. These results can be considered as inconsistent, but the LC model is known to produce somewhat small PI (Keilman and Pham 2006). As mentioned previously, the wider PI from the CoDa models might come from the more pronounced fluctuations of the CoDa time index. The random variation of the respective mortality matrices seems to be captured by $\kappa_t$ with CoDa, but by $\beta_x$ with the LC model, as shown in Figures 1 and 2. Future research should try to provide a more detailed explanation for these results, look deeper into the causes of the wider PI for CoDa, and consider new ways to estimate the PI.

As mentioned previously, both $k_{t,i}$ in the LC-coherent and CoDa-coherent are not guaranteed to reach a constant, e.g., if the trend has recorded a long-term increasing or decreasing trend. In this case, the coherent model might fail, as the population's mortality diverges more and more from the average trend. Mortality for such populations should perhaps not be forecast coherently in the remaining countries. However, in our results such patterns were rarely observed for the selected countries.

In this study, we applied the same methodology to all Western European countries and presented the model to make as valid a comparison as possible with the original LC model (Lee and Carter 1992) and its coherent extension (Li and Lee 2005). We did so to show the potential of the method as a general forecasting model and show its adaptability with commonly used models. This paper is a first attempt to explore the use of CoDa in a coherent forecasting context and has shown that CoDa overcomes some

shortcomings of the LC model. Oeppen (2008) also showed that this method provides interesting possibilities for forecasting mortality by causes of death.

Further development of the method in different contexts should be the subject of future research, including, among others, cohort forecasts. By reading our results in a cohort perspective, the life expectancy at birth for females born in 1960 in France is predicted to be 88.8 years with the LC-coherent model and 93.1 with the CoDa-coherent model. However, no information on cohort effects has been considered in the models to produce proper cohort forecasts. This could be attained, for example, by adapting the Renshaw and Haberman (2006) model to CoDa.

# 7  Conclusion

Both LC and CoDa models and their coherent variants (LC-coherent and CoDa-coherent) share some similarities: the parameters $\beta_x$ and $\kappa_t$ are found by applying a SVD to a centered matrix and the time index is extrapolated using time series models. However, the models differ in many ways. The key difference is that the forecasts are based on different indicators: $m_{t,x}$ and $d_{t,x}$. The use of a specific indicator implies a method adapted to the indicator's characteristics, as presented earlier. The use of different indicators and methods implies that the parameters have different interpretations and the models have different assumptions. As mentioned previously, the CoDa model is not based on a constant RMI assumption, as with the LC model. Furthermore, over time, the $d_{t,x}$ are not free to vary independently from one another, as the $m_{t,x}$ can, an aspect which appears in their covariance structure.

In this article, we forecast mortality acknowledging that there is coherence among Western European countries using compositional data analysis of life table deaths. This procedure is a promising new way to provide a coherent mortality forecast, as it 1) preserves coherence among countries, 2) acknowledges covariance between components, 3) explains a large proportion of the observed variability, and 4) allows the rate of mortality improvement to change over time. Our results show that using a CoDa-coherent model to forecast mortality for the period 1995–2011 increased the accuracy of the forecast for many of the selected countries.

# 8  Acknowledgement

# References

Aitchison, J. (1986). *The statistical analysis of compositional data*. London: Chapman and Hall.

Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J., and Pawlowsky-Glahn, V. (2000). Logratio analysis and compositional distance. *Mathematical Geology* 32(3): 271–275. doi:10.1023/A:1007529726302.

Barbieri, M., Wilmoth, J., Shkolnikov, V., Glei, D., Jasilionis, D., Jdanov, D., Boe, C., Riffe, T., Grigoriev, P., and Winant, C. (2015). Data resource profile: The Human Mortality Database (HMD). *International Journal of Epidemiology* 44(5): 1549–1556. doi:10.1093/ije/dyv105.

Bergeron-Boucher, M.P., Ebeling, M., and Canudas-Romo, V. (2015). Decomposing changes in life expectancy: Compression versus shifting mortality. *Demographic Research* 33(14): 391–424. doi:10.4054/DemRes.2015.33.14.

Bohk-Ewald, C. and Rau, R. (2017). Probabilistic mortality forecasting with varying age-specific survival improvements. *Genus* 73(1): 1–37. doi:10.1186/s41118-016-0017-8.

Booth, H., Hyndman, R., Tickle, L., and de Jong, P. (2006). Lee–Carter mortality forecasting: A multi–country comparison of variants and extensions. *Demographic Research* 15(9): 289–310. doi:10.4054/DemRes.2006.15.9.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee–Carter under conditions of variable mortality decline. *Population Studies* 56(3): 325–336. doi:10.1080/00324720215935.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science* 3(1–2): 3–43. doi:10.1017/S1748499500000440.

Brass, W. (1971). On the scale of mortality. In: Brass, W. (ed.). *Biological aspects of demography*. London: Taylor and Francis: 69–110.

Cairns, A.J., Blake, D., Dowd, K., Coughlan, G.D., and Khalaf-Allah, M. (2011). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin* 41(1): 29–59. doi:10.2143/AST.41.1.2084385.

Camarda, C. (2012). Mortality smooth: An R package for smoothing poisson counts with p-splines. *Journal of Statistical Software* 50(1): 1–24. doi:10.18637/jss.v050.i01.

Carter, L.R. and Lee, R.D. (1992). Modeling and forecasting US sex differentials in mortality. *International Journal of Forecasting* 8(3): 393–411. doi:10.1016/0169-2070(92)90055-E.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology* 35(3): 279–300. doi:10.1023/A:1023818214614.

HMD (2016). Human Mortality Database. [electronic resource]. University of California, Berkeley and Max Planck Institute for Demographic Research. www.mortality.org.

Hyndman, R.J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography* 50(1): 261–283. doi:10.1007/s13524-012-0145-5.

Kannisto, V., Lauritsen, J., Thatcher, A.R., and Vaupel, J.W. (1994). Reductions in mortality at advanced ages: Several decades of evidence from 27 countries. *Population and Development Review* 20(4): 793–810.

Keilman, N. and Pham, D.Q. (2006). *Prediction intervals for Lee–Carter-based mortality forecasts*. Paper presented at the European Population Conference 2006, Liverpool, United Kingdom, June 21–24, 2006.

Lee, R.D. (1998). Probabilistic approaches to population forecasting. *Population and Development Review* 24(419): 156–190. doi:10.2307/2808055.

Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association* 87(419): 659–671.

Lee, R.D. and Miller, T. (2001). Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography* 38(4): 537–549. doi:10.1353/dem.2001.0036.

Li, N. and Lee, R.D. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography* 42(3): 575–594. doi:10.1353/dem.2005.0021.

Lloyd, C.D., Pawlowsky-Glahn, V., and Egozcue, J.J. (2012). Compositional data analysis in population studies. *Annals of the Association of American Geographers* 102(6): 1251–1266. doi:10.1080/00045608.2011.652855.

Martín-Fernández, J., Barceló-Vidal, C., and Pawlowsky-Glahn, V. (2003). Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology* 35(3): 253–278. doi:10.1023/A:1023866030544.

Mert, M.C., Filzmoser, P., Endel, G., and Wilbacher, I. (2016). Compositional data analysis in epidemiology. *Statistical Methods in Medical Research* 1–14. doi:10.1177/0962280216671536.

Oeppen, J. (2006). Life expectancy convergence among nations since 1820: Separating the effects of technology and income. In: Bengtsson, T. (ed.). *Perspectives on mortality forecasting III. The linear rise in life expectancy: History and prospects*. Stockholm,

Swedish Social Insurance Agency: 55–82.

Oeppen, J. (2008). *Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data*. Paper presented at the European Population Conference 2008, Barcelona, Spain, July 9–12, 2008.

Oeppen, J. and Vaupel, J.W. (2002). Broken limits to life expectancy. *Science* 296(5570): 1029–1031. doi:10.1126/science.1069675.

Pawlowsky-Glahn, V. and Buccianti, A. (2011). *Compositional data analysis: Theory and applications*. New Jersey: John Wiley. http://www.dx.doi.org/10.1002/9781119976462 doi:10.1002/9781119976462.

Pawlowsky-Glahn, V. and Egozcue, J. (2006). Compositional data and their analysis: An introduction. *Geological Society* 264(1): 1–10. doi:10.1144/GSL.SP.2006.264.01.01.

Preston, S., Heuveline, P., and Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Oxford: Blackwell.

Raftery, A.E., Chunn, J.L., Gerland, P., and Ševčíková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography* 50(3): 777–801. doi:10.1007/s13524-012-0193-x.

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

Russolillo, M., Giordano, G., and Haberman, S. (2011). Extending the Lee–Carter model: A three-way decomposition. *Scandinavian Actuarial Journal* 2011(2): 96–117. doi:10.1080/03461231003611933.

Schinzinger, E., Denuit, M.M., and Christiansen, M.C. (2016). A multivariate evolutionary credibility model for mortality improvement rates. *Insurance: Mathematics and Economics* 69(2): 70–81. doi:10.1016/j.insmatheco.2016.04.004.

Stoeldraijer, L., van Duin, C., van Wissen, L., and Janssen, F. (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. doi:10.4054/DemRes.2013.29.13.

Thatcher, R.A., Kannisto, V., and Vaupel, J.W. (1998). *The force of mortality at ages 80 to 120*. Odense: Odense University Press.

Torri, T. and Vaupel, J. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting* 28(2): 519–531. doi:10.1016/j.ijforecast.2011.01.009.

Vaupel, J. and Yashin, A. (1987). Repeated resuscitation: How lifesaving alters life tables. *Demography* 24(1): 123–135. doi:10.2307/2061512.

White, K.M. (2002). Longevity advances in high-income countries, 1955–1996. *Population and Development Review* 28(1): 59–76. doi:10.1111/j.1728-4457.2002.00059.x.

Wilmoth, J.R. (1995). Are mortality projections always more pessimistic when disaggregated by cause of death? *Mathematical Population Studies* 5(4): 293–319. doi:10.1080/08898489509525409.

Wilson, C. (2001). On the scale of global demographic convergence 1950–2000. *Population and Development Review* 27(1): 155–171. doi:10.1111/j.1728-4457.2001.00155.x.

Wilson, C. (2011). Understanding global demographic convergence since 1950. *Population and Development Review* 37(2): 375–388. doi:10.1111/j.1728-4457.2011.00415.x.

## Appendix A: CoDa operators and method

Table A-1 summarizes the different CoDa operators and concepts used in this paper.

**Table A-1:** **CoDa operators and methods used, their descriptions, and equations**

| CoDa operator | Description | Equation |
|---|---|---|
| Composition | Vector of $c$ components representing part of a whole and summing up to a constant $K$. | $X = [x_1, x_2, ..., x_c]$ $\sum\limits_{i=1}^{c} x_i = K$ |
| $C[]$ | This procedure is called closing. To close the data following certain operations and transformations on a composition, e.g., $Y = X^2$, the proportions of the vector are calculated and then multiplied by the constant sum chosen. This procedure ensures that the estimates in compositional data sum up to the initial constant. | $C[Y] = \left[\frac{y_1}{\sum y_i}, \frac{y_2}{\sum y_i}, ..., \frac{y_c}{\sum y_i}\right] \cdot K$ |
| $\oplus$ | Standard operation in compositional data analysis named perturbation. To perturb a composition $X$ by another composition $Y$, calculate the component-wise product and then close the result. | $Z = X \oplus Y =$ $C[x_1 y_1, x_2 y_2, ..., x_c y_c]$ |
| $\ominus$ | Standard operation in compositional data analysis consisting in perturbing a composition by the inverse element of another composition. To come back to $X$, divide $Z$ component-wise by $Y$ and then close the result. $\ominus Y$ is named the inverse element of $Y$. | $Z \ominus Y =$ $C\left[\frac{z_1}{y_1}, \frac{z_2}{y_2}, ..., \frac{z_c}{y_c}\right] = X$ |

**Table A-1:**     **(Continued)**

| CoDa operator | Description | Equation |
|---|---|---|
| $clr()$ | The centered log-ratio ($clr$) transformation is one of the log-ratio representations of compositional data. This transformation is used to represent a composition as a real vector ($U$), on which standard statistical analyses can be used. The $clr$-coordinates of a vector $X$ are the logarithm of the components divided by its geometric mean. | $g = (x_1 \cdot x_2 \cdot ... \cdot x_c)^{1/c}$ $clr(X) = [ln(\frac{x_1}{g}), ln(\frac{x_2}{g}), ..., ln(\frac{x_c}{g})]$ $clr(X) = U$ |
| $clr^{-1}()$ | The inverse $clr$ transformation is the procedure used to re-enter compositional data form, following a $clr$ transformation (from $U$ to $X$). The exponential of $clr$-coordinates are obtained and then closed. | $clr^{-1}(U) = C[e^{u_1}, e^{u_2}, ..., e^{u_c}]$ $clr^{-1}(U) = X$ |
| $AD$ | The Aitchison distance is a measure of dissimilarity between two compositions. In this paper, the AD measure is used as a measure of accuracy between a forecast composition and an observed composition. The AD is the square root of the sum of the squared difference between two compositions expressed in $clr$-coordinates. | $AD = \big[ \sum\limits_{i=1}^{c} (clr(x_i) - clr(y_i))^2 \big]^{1/2}$ |

*Source*: Aitchison 1986; Pawlowsky-Glahn and Egozcue 2006; Pawlowsky-Glahn and Buccianti 2011

Additionally, we give here the step-by-step procedure of the CoDa method presented in the main text to forecast $d_{t,x}$:

1. We start from a matrix $D$ of the life table deaths ($d_{t,x}$), with $T$ rows representing the number of years and $X + 1$ columns representing the ages $x$. This follow the CoDa notational convention that each row represents a composition (Pawlowsky-Glahn and Egozcue 2006). The sum of each row adds up to the life table radix.

2. We obtain a second matrix $F$, with elements $f_{t,x}$, by perturbing the matrix $D$ by the column geometric means for each age, $\alpha_x$, by using the CoDa perturbation operator $\ominus$. This step centers the matrix to better visualize the structure:

$$f_{t,x} = d_{t,x} \ominus \alpha_x.$$

3. The next step is to unrestrict the data. Aitchison (1986) showed that compositional data is confined to a restricted space where the components can only vary between 0 and a given limit. Aitchison (1986) suggested using log-ratio transformations to allow the data to vary freely. We here apply the centered log-ratio ($clr$) transformation:

$$h_{t,x} = clr(f_{t,x}) = ln(\frac{f_{t,x}}{g_t}), \tag{9}$$

where $g_t$ are the geometric means over age at time $t$. We thus obtain a new transformed matrix, $H$, with elements $h_{t,x}$. This new space, where the data is free to vary from $-\infty$ to $\infty$, is known as the "real space."

4. A singular value decomposition (SVD) is then applied to the matrix $H$.

5. A low-rank approximation of the matrix $H$, $H^*$, is constructed and forecast. Oeppen (2008) compared a rank-1 and rank-2 approximation of the matrix $H$ and selected an ARIMA(0,2,2) model to forecast the time index for Japan, based on the best AIC criterion. In this article, we suggest using a rank-1 approximation of the matrix $H$, as no major gains in explained variance are obtained, for most countries, by using a rank-2 approximation. The time index is forecast with an ARIMA(0,1,1) model, based on the best BIC value.

6. To transform the matrix back into compositional data, $F^*$, the inverse centered log-ratio is used:

$$f_{t,x}^* = clr^{-1}(h_{t,x}^*) = C[e^{h_{t,x}^*}], \tag{10}$$

where $h_{t,x}^*$ are the elements of the matrix $H^*$ and $C[]$ is a closing procedure (see Table A-1).

7. The last step is to compositionally add back the geometric means, to obtain the matrix $D^*$:

$$d_{t,x}^* = f_{t,x}^* \oplus \alpha_x. \tag{11}$$

## Appendix B: Prediction intervals

The method used to calculate the prediction intervals (PI) is built on the Keilman and Pham (2006) model and allows us to consider two sources of uncertainty in the forecasts: Estimates of the parameters and extrapolated values of the time index. The following steps apply to the PI of the LC and CoDa models.

1. Estimate the model, extrapolate the time index $\kappa_t$ using the selected time series model, and construct the matrix $\kappa_t \beta_x$.

2. For each time-age interval, the model provides an error ($\epsilon_{t,x}$) (Keilman and Pham 2006). For the CoDa model, the errors are found within the $clr$ transform (see equation (5)). We thus make the hypothesis that the parameter $\alpha_x$, found before the $clr$ transformation, is correct. The residuals ($\epsilon_{t,x}$) are placed in a table. A new table of residuals is created by assigning to each age and year a randomly chosen row and column of the original table. The simulated residuals are added to the fitted value, $\kappa_t \beta_x$ in step 1. This random allocation procedure is repeated $n_\epsilon$ times. For each of the $n_\epsilon$ new tables of values, the model chosen in step 1 is estimated. We thus obtain $n_\epsilon$ estimates of $\kappa_t$ and $\beta_x$, and take into account the uncertainty in estimating these parameters.

3. For each simulation of step 2, a new time index is found and extrapolated using the selected time series model. At each $n_\epsilon$ estimate of $\kappa_t$, PI are estimated using $n_\kappa$ simulations with resampled errors (bootstrap). We obtain a set of $n_\epsilon n_\kappa$ future mortality trends. This step considers the uncertainty in the extrapolated value of the time index.

4. For each of the $n_\epsilon n_\kappa$ future mortality trends, a life table is calculated. Prediction intervals for age-specific death rates, life table deaths, and life expectancy are obtained by finding the 0.025 and 0.975 percentiles of the simulated data for the 95% PI and the 0.1 and 0.9 percentiles for the 80% PI.

With the LC-coherent and CoDa-coherent models, the uncertainty in the model comes from the common factor $\kappa_t \beta_x$ and from the deviation factor $k_{t,i} b_{x,i}$. Assuming independence between these factors, the PI can be found by applying the previous step

for both common and deviation factors, obtaining $n_\epsilon n_\kappa$ simulations for the common factor and $n_\epsilon^d n_\kappa^d$ simulation for the deviation term. The errors for the deviation factor at step 2 are found by $clr(d_{t,x} \ominus \alpha_x \ominus e^{\kappa_t \beta_x}) = k_{t,i} b_{x,i} + \epsilon_{t,x,i}$. For each $n_\epsilon n_\kappa$ simulation of the common factor, we added the $n_\epsilon^d n_\kappa^d$ simulation for the deviation. We thus obtain $n_\epsilon n_\kappa n_\epsilon^d n_\kappa^d$ sets of future mortality trends. The number of simulations are $n_\epsilon$=100, $n_\kappa$=100, $n_\epsilon^d$=100, and $n_\kappa^d$=100, leading to a total of 100,000,000 simulations.

# Appendix C: Data

## Germany

Data for Germany in the HMD is available starting in 1990 only. However, data is available for East and West Germany separately starting in 1956. To obtain longer series for Germany, we combined death counts and exposure to risk data for East and West Germany, taking account of their population size. Life tables for Germany were then calculated starting in 1960.

## Average mortality

The average mortality for the 15 selected countries is based on the average of the observed age-specific death rates ($\bar{m}_{t,x}$):

$$\bar{m}_{t,x} = \frac{\sum\limits_{i=1}^{I} m_{t,x,i}}{I},$$

where $I$ is the number of countries. These average age-specific death rates weight all the countries equally irrespective of their population size. A life table is then calculated using the average death rates following standard methods (Preston, Heuveline, and Guillot 2001).

## Age 80 to 120 smoothed with a Kannisto model

For many low mortality countries, extrapolating past trends tends to shift the density distribution of deaths to higher ages and to approach the last age available in the HMD; i.e., 110. To avoid an artificial compression against this arbitrary limit, we extend mortality trends until age 120. To do so, we used the Kannisto model (Thatcher, Kannisto, and

Vaupel 1998) for old-age mortality and applied it to ages 80 to 120 using a Poisson log-likelihood procedure.

### Problems with zeros

When zeros are present in a composition, the log-ratio representation of compositional data is problematic (Martín-Fernández, Barceló-Vidal, and Pawlowsky-Glahn 2003). By applying a Kannisto model to ages 80 to 120, we avoid the problem at old ages. However, for some countries, life table deaths can equal 0 at younger ages for some specific years.

In a life table context, the 0 values occur because no deaths have been observed or counted at a specific age $x$ and time $t$. Treatment of 0 values is thus done on the observed death counts ($D_{t,x}$). Procedures were suggested by Martín-Fernández, Barceló-Vidal, and Pawlowsky-Glahn (2003) to treat zero counts (essential zeros). We use a multiplicative replacement strategy. If we have a composition $X$ of the observed deaths $D_x$ with P parts, $X = [x_1, x_2, ...., x_P]$, containing zeros, we want to replace it by a composition R with P parts, $R = [r_1, r_2, ...., r_P]$, without zeros:

$$r_j = \begin{cases} \delta, & \text{if } x_j = 0 \\ (1 - \frac{z\delta}{K})x_j, & \text{if } x_j > 0 \end{cases}$$

where $\delta$ is the imputed value on part $x_j$, $z$ is the number of zeros counted in the composition $X$, and $K$ is the constant of the sum constraint ($\sum x_j = K$). The value of $\delta$ is half of the minimum $D_{t,x}$ observed over all ages and years, when $D_{t,x} > 0$, divided by the total number of deaths observed the year the zero was recorded:

$$\delta_t = \frac{\min\limits_{t,x}(D_{t,x})/2}{\sum\limits_{x=0}^{120} D_{t,x}} \quad \forall D_{t,x} > 0.$$

Once the composition R is found, we multiply it by $\sum\limits_{x=0}^{120} D_{t,x}$ to create a new set of death counts without zeros. The death rates are calculated based on these last death counts without zeros.

## Appendix D: The fitted models

Figure A-1 shows the life table deaths for Spanish females, on a log scale, at selected ages (0, 15, 30, 45, 60, 75, 90, and 105) observed and fitted with the LC, LC-coherent, CoDa, and CoDa-coherent models. For most age groups and for all four models, the fit is generally good. The coefficient of determination ($R^2$) value for ages 0, 45, 60, 75, 90, and 105 is 90% and over for all four models. The fit at ages 15 and 30 is however poorer, especially for the CoDa and CoDa-coherent models, with $R^2$ values between 75% and 90%. This value is over 85% with the LC and LC-coherent models at these same two ages. The number of deaths at age 15 and 30 are, however, relatively low, and the errors in modeling and forecasting them will thus have little impact on life expectancy (Lee and Carter 1992). On the other hand, both CoDa models fit the mortality at higher ages very well. The number of deaths at these ages is often important and has been more influential on life expectancy since the second half of the 20th century (Bergeron-Boucher, Ebeling, and Canudas-Romo 2015).

Using the coherent version of the LC and CoDa models also moderately increases the fit at some ages. Similar results are found when looking at the model fits for the $m_{t,x}$.

**Figure A-1:**   **Life table deaths ($d_{t,x}$) at specific ages (0, 15, 30, 45, 60, 75, 90, and 105) with a radix of 1 observed (dot) and fitted (lines) with the LC and CoDa models, as well as their coherent extension, Spanish females, 1960–2011**
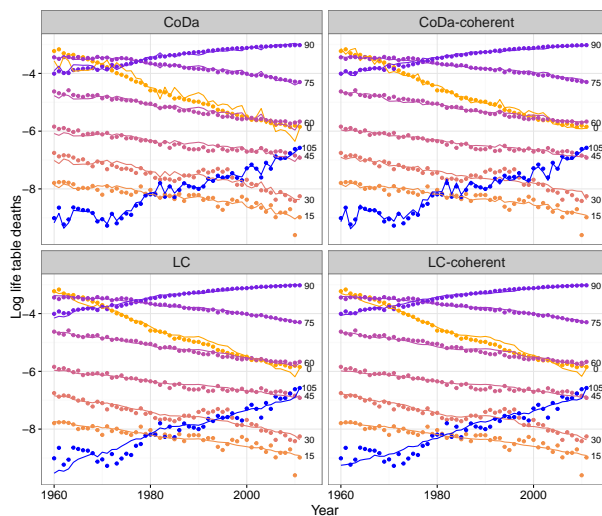


*Source*: HMD (2016) and authors' calculations.

## Appendix E: Evaluating the models, all countries

**Figure A-2:** **Female life expectancy at birth observed from 1960 to 2011 (in black) and forecast from 1995 to 2011 for 15 European countries using four forecasting models**
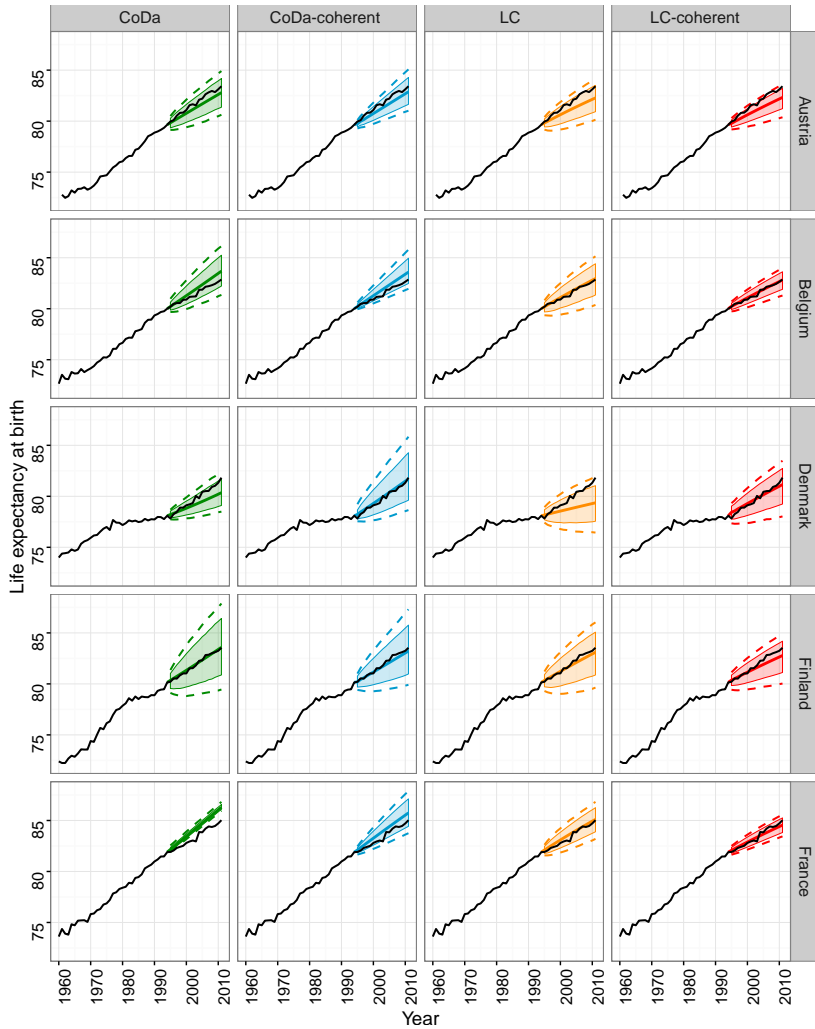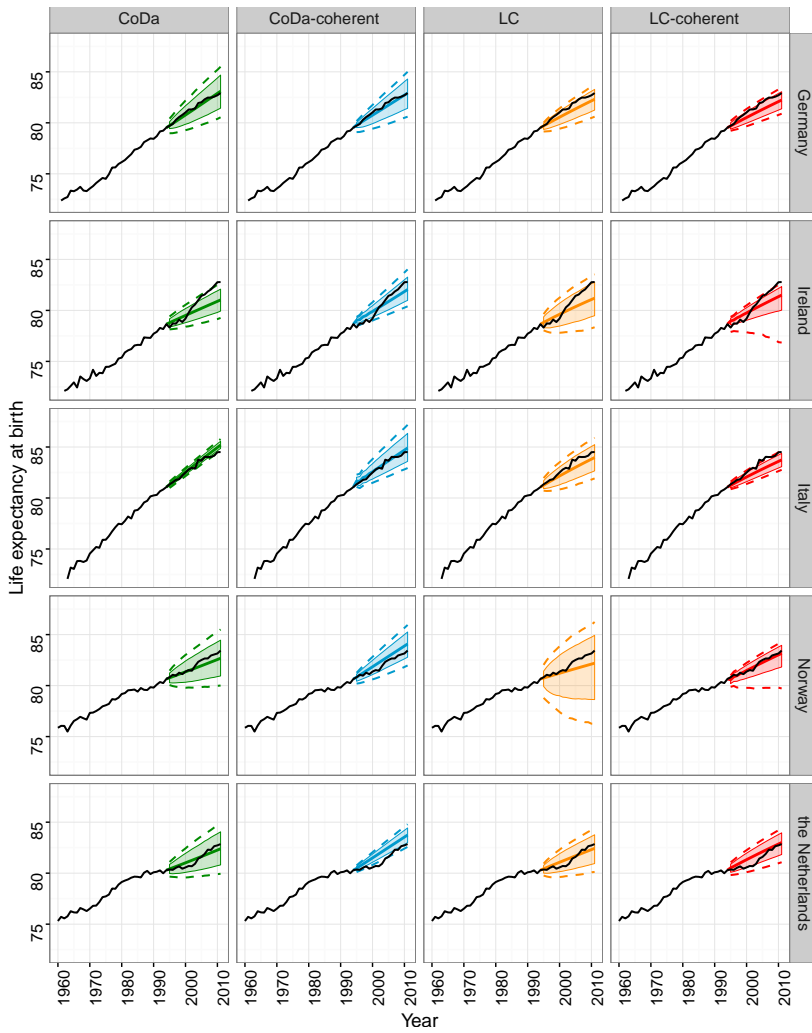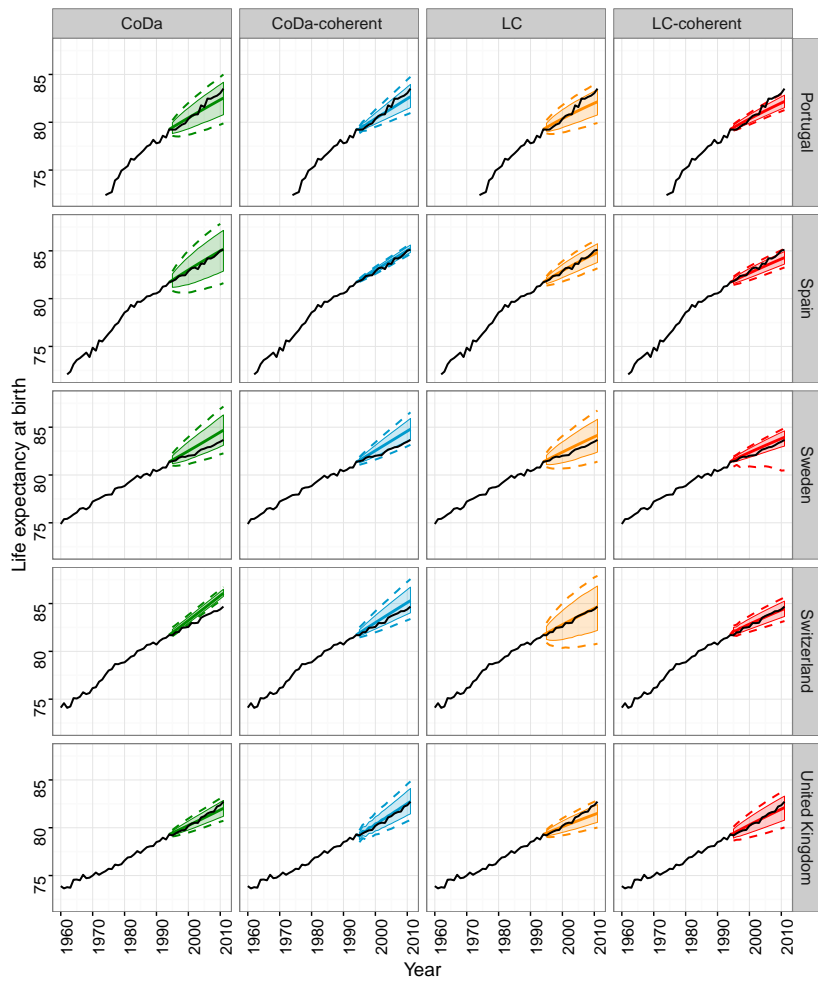
**Figure A-2:**     **(Continued)**

**Figure A-2:** (Continued)



*Source*: HMD (2016) and authors' calculations.

## Appendix F: Rates of mortality improvement (RMIs)

The rate of mortality improvement (RMI) at each age implied by the Lee–Carter (LC) model is not supported by empirical findings (e.g., Kannisto et al. 1994). The LC model assumes constant RMIs, while empirical data shows that the RMIs have been increasing, especially at older ages (Kannisto et al. 1994). When using a CoDa model, the RMIs can change over time. The RMI for an indicator $I$ forecast with a model $M$ is here defined by:

$$RMI_{t,x}^{I,M} = -\frac{\dot{I}_{t,x}}{I_{t,x}}, \tag{12}$$

where the dot on the top of the variable indicates its derivative with respect to time. For the Lee–Carter model, the RMI calculated for the $m_{t,x}$ is equal to

$$RMI_{t,x}^{m,LC} = -\dot{\kappa}_t \beta_x, \tag{13}$$

where $\dot{\kappa}_t$ is equal to the drift when forecasting with a random walk with drift: $\dot{\kappa}_t = d + \epsilon_t; \epsilon_t = 0$. The RMIs for the LC model is thus constant over time, although differing from age to age. When the life table radix is 1, the CoDa model can be rewritten as:

$$\hat{d}_{t,x} = \alpha_x \frac{e^{\beta_x \kappa_t}}{S_{clr,t}} \frac{1}{S_{\alpha_t}}, \tag{14}$$

where $S_{\alpha,t}$ and $S_{clr,t}$ are the sum at time $t$ of the matrices $\alpha_x C[e^{\kappa_t \beta_x}]$ and $e^{\kappa_t \beta_x}$ respectively, used in the closure procedure, as

$$C[e^{\kappa_t \beta_x}] = \frac{e^{\kappa_t \beta_x}}{S_{clr,t}}$$

and

$$\alpha_x \oplus C[e^{\kappa_t \beta_x}] = \frac{\alpha_x C[e^{\kappa_t \beta_x}]}{S_{\alpha,t}}.$$

From equation (14), the RMI for the $d_{t,x}$ with the CoDa model can be derived and is equal to:

$$RMI_{t,x}^{d,CoDa} = \frac{\dot{S}_{\alpha,t}}{S_{\alpha,t}} + \frac{\dot{S}_{clr,t}}{S_{clr,t}} - \dot{\kappa}_t\beta_x. \qquad (15)$$

As for the LC model, $\dot{\kappa}_t$ with the CoDa model is equal to the drift when forecasting with a random walk with drift, making the term $-\dot{\kappa}_t\beta_x$ a constant. Thus, the terms $\frac{\dot{S}_{\alpha,t}}{S_{\alpha,t}}$ and $\frac{\dot{S}_{clr,t}}{S_{clr,t}}$ determine if $RMI_{t,x}^{d,CoDa}$ is constant or not. To assess how $RMI_{t,x}^{d,CoDa}$ changes over time, we calculated the RMIs for the forecast $\hat{d}_{t,x}$, using a random walk with drift to forecast $\kappa_t$. The RMIs for discrete data can be estimated as:

$$RMI_{t,x}^{I,M} = -ln(\frac{I_{t+1,x}}{I_{t,x}}). \qquad (16)$$

Figure A-3 shows that the RMIs for the $\hat{d}_{t,x}$ forecast with CoDa are not constant over time. The increase of the $RMI_{t,x}^{d,CoDa}$ over time is not linear: The increase is accelerating until the middle of the 2030s and then starts to decelerate. The RMIs at each age evolve in parallel and the difference between two consecutive ages is equal to $-\dot{\kappa}_t(\beta_x - \beta_{x+1})$. At some ages, the RMI is negative, e.g., 105, meaning that the density of deaths is increasing at these ages.

**Figure A-3:** **Rate of mortality improvement at specific ages for French females' life table deaths ($d_{t,x}$) forecast with a CoDa methodology, 2011–2050**
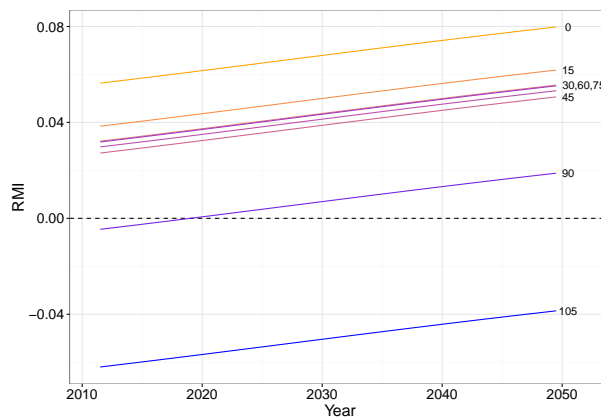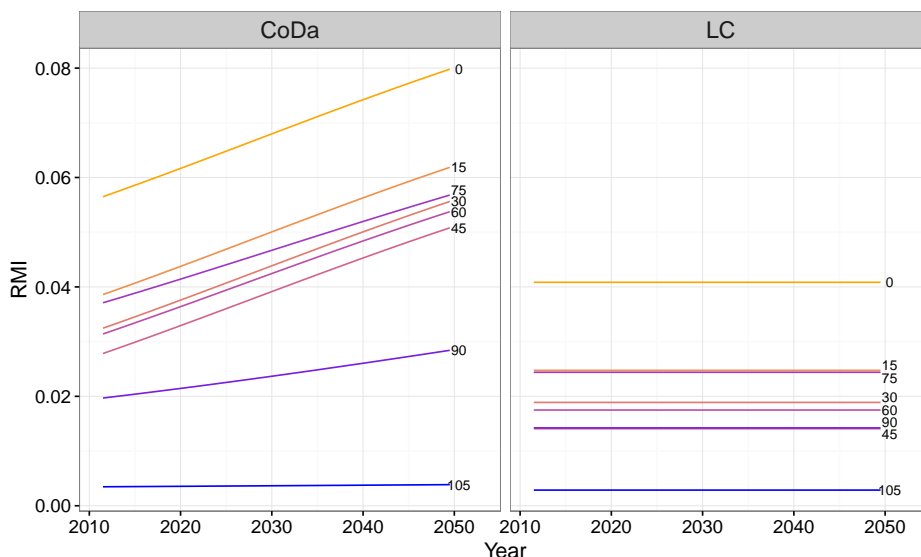


*Source*: HMD (2016) and authors' calculations.

The RMIs for two different indicators are hard to compare. Thus, from the forecasts of CoDa based on $d_{t,x}$, age-specific death rates ($m_{t,x}$) were constructed and their RMIs calculated. Figure A-4 shows the RMIs of the $m_{t,x}$ from both LC and CoDa models. The figure confirms that the RMIs for the CoDa model are not constant at all ages.

**Figure A-4:** **Rate of mortality improvement at specific ages for French females' death rates ($m_{t,x}$) forecast with an LC and CoDa methodology, 2011–2050**



*Source*: HMD (2016) and authors' calculations.

## Appendix G: Jump-off adjustment

We adjusted the forecasts to correct for the jump-off year level for the CoDa and CoDa-coherent models using the following equation:

$$d^j_{T:T+N,x} = \hat{d}_{T:T+N,x} \oplus [d_{T,x} \ominus \hat{d}_{T,x}], \tag{17}$$

where $T$ is the last year observed, $N$ is the numbers of years forecast and $d_{t,x}$, $\hat{d}_{t,x}$ and $d^j_{t,x}$ are the life table deaths observed, fitted, and forecast, and adjusted for the jump-off level, respectively. We use a similar method for the LC and LC-coherent models:

$$ln(m^j_{T:T+N,x}) = ln(\hat{m}_{T:T+N,x}) + [ln(m_{T,x}) - ln(\hat{m}_{T,x})], \tag{18}$$

where $m_{t,x}$ are the age-specific death rates at time $t$. To avoid extrapolating the random variation of the last year observed ($T$), we smooth it using a P-spline smoothing procedure for Poisson death counts (Camarda 2012).

We also adjusted the PI in such a way that the median of the simulations, used to calculate the PI, is equal to the forecast value for the life expectancy at birth:

$$B^j_{0,t} = B_{0,t} + (\hat{e}_{0,t} - M_{0,t}), \tag{19}$$

where $B_{0,t}$ is the PI bounds (upper or lower) of the life expectancy at birth at time $t$ and $\hat{e}_{0,t}$ and $M_{0,t}$ are the life expectancy at birth and the median forecasts, respectively. For most cases, the median was very close to the forecast life expectancy.

Furthermore, the forecast of $\kappa_t$, in some cases, recorded a break in its trend at year $T + 1$ when forecasting with the ARIMA(0,1,1) model due to the MA component. We thus also adjust $\kappa_t$ such as:

$$\kappa^j_{T+1:T+n} = \kappa_{T+1:T+n} + [d - (\kappa_{T+1} - \kappa_T)] \tag{20}$$

where $d$ is the drift of the ARIMA(0,1,1) model with drift used for the CoDa model.