



Heriot-Watt University  
Research Gateway

## A multivariate evolutionary credibility model for mortality improvement rates

### Citation for published version:

Schinzinger, E, Denuit, MM & Christiansen, M 2016, 'A multivariate evolutionary credibility model for mortality improvement rates', *Insurance: Mathematics and Economics*, vol. 69, pp. 70-81.  
<https://doi.org/10.1016/j.insmatheco.2016.04.004>

### Digital Object Identifier (DOI):

[10.1016/j.insmatheco.2016.04.004](https://doi.org/10.1016/j.insmatheco.2016.04.004)

### Link:

[Link to publication record in Heriot-Watt Research Portal](#)

### Document Version:

Early version, also known as pre-print

### Published In:

Insurance: Mathematics and Economics

### General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [open.access@hw.ac.uk](mailto:open.access@hw.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# A MULTIVARIATE EVOLUTIONARY CREDIBILITY MODEL FOR MORTALITY IMPROVEMENT RATES

EDO SCHINZINGER\*  
DFG Research Training Group 1100  
University of Ulm  
Ulm, Germany

\*Corresponding author: [edo.schinzinger@uni-ulm.de](mailto:edo.schinzinger@uni-ulm.de)

MICHEL M. DENUIT  
Institut de Statistique, Biostatistique et Sciences Actuarielles  
Université Catholique de Louvain  
Louvain-la-Neuve, Belgium

MARCUS C. CHRISTIANSEN  
Department of Actuarial Mathematics and Statistics  
School of Mathematical and Computer Sciences  
Heriot-Watt University  
Edinburgh, United Kingdom

June 21, 2015

## **Abstract**

The present paper proposes an evolutionary credibility model that describes the joint dynamics of mortality through time in several populations. Instead of modeling the mortality rate levels, the time series of population-specific mortality rate changes, or mortality improvement rates are considered and expressed in terms of correlated time factors, up to an error term. Dynamic random effects ensure the necessary smoothing across time, as well as the learning effect. They also serve to stabilize successive mortality projection outputs, avoiding dramatic changes from one year to the next. Statistical inference is based on maximum likelihood, properly recognizing the random, hidden nature of underlying time factors. Empirical illustrations demonstrate the practical interest of the approach proposed in the present paper.

*Key words and phrases:* mortality projection; predictive distribution; multi-population modeling.

# 1 Introduction

Mortality forecasts are used in a wide variety of fields. Let us mention health policy making, pharmaceutical research, social security, retirement fund planning and life insurance, to name just a few.

Following the elegant approach to mortality forecasting pioneered by Lee and Carter (1992) many projection models decompose the death rates (on the logarithmic scale) or the one-year death probabilities (on the logit scale) into a linear combination of a limited number of time factors. See, e.g., Hunt and Blake (2014). In a first step, regression techniques are used to extract the time factors from the available mortality data. In a second step, the time factors are intrinsically viewed as forming a time series to be projected to the future. The actual age-specific death rates are then derived from this forecast using the estimated age effects. This in turn yields projected life expectancies.

In the first step of the two-step model calibration procedure, the random nature of the unobservable time factor is disregarded, and this may bias the analysis. As possible incoherence may arise from this two-step procedure, Czado et al. (2005) integrated both steps into a Bayesian version of the model developed by Lee and Carter (1992) in order to avoid this deficiency. After Czado et al. (2005), Pedroza (2006) formulated the Lee-Carter method as a state-space model, using Gaussian error terms and a random walk with drift for the mortality index. See also Girosi and King (2008), Kogure et al. (2009), Kogure and Kurachi (2010) and Li (2014) for related works. However, the practical implementation of Bayesian methods often requires computer-intensive Markov Chain Monte Carlo (MCMC) simulations. This is why we propose in this paper a simple credibility model ensuring robustness over time while keeping the computational issues relatively easy and allowing for the flexibility of time series modeling. It is worth stressing that the time factor is here treated as such, and not as a parameter to be estimated from past mortality statistics using regression techniques before entering time series models. In this way, we recognize the hidden nature of the time factor and its intrinsic randomness.

Whereas most mortality studies consider both genders separately, the model proposed in this paper easily combines male and female mortality statistics. This is particularly useful in practice when both genders are usually involved. In insurance applications, for instance, separate analyses could lead to miss this strong dependence pattern, which considerably reduces possible diversification effects between male and female policyholders inside the portfolio. In demographic projections, combining male and female data is necessary to ensure consistency in gender-specific mortality forecast. This problem has been considered by several authors in the literature. Let us mention Carter and Lee (1992) who fitted the Lee and Carter (1992) model to male and female populations separately and then measured the dependence between the two gender-specific time factors. These authors considered three models for the pair of time factors: a bivariate random walk with drift, a single time factor common to both genders and a co-integrated process where the male index follows a random walk with drift and there exists a stationary linear combination of both time factors.

More generally, the credibility model proposed in this paper is able to pool several populations to produce mortality forecasts for a group of countries. In such a context, Yang and Wang (2013) assumed that the time factors followed a vector error correction model. See also Zhou et al. (2013). Other models incorporate a common factor for the combined population as a whole, as well as additional factors for each sub-population. The common factor describes the main long-term trend in mortality change while the additional factors depict the short-term discrepancy from the main trend inside each sub-population. See Li and Lee (2005) who proposed the augmented common factor model generalized by Li (2013) to several factors. The model structure proposed in Delwarde et al. (2006), by Debón et al. (2011), and by Russolillo et al. (2011) only include a single, common time factor. As argued in Carter and Lee (1992), this simple arrangement may enforce greater consistency and is a parsimonious way to model both populations jointly. However, it also implies that the death rates of the two populations are perfectly associated, an assumption with far-reaching consequences in risk management.

Our paper innovates in that the new multi-population mortality projection model we propose is based on mortality improvement rates instead of levels. Recently, several authors suggested to target improvement rates to forecast future mortality, instead of the death rates. While the time dependence structure of death rate models are dominated by the continuing downward trend, the improvement rates are already trend adjusted. See, e.g., Mitchell et al. (2013) or Aleksic and Börger (2011). The model developed in this paper appears to be useful for studying securitization mechanisms, as shown by the Kortis bond issued by Swiss Re in 2010. The payoff of this first longevity trend bond is linked to the divergence in mortality improvement rates between two countries (US and UK) and thus nicely fits our proposed model.

Furthermore, the model is fitted properly, recognizing the hidden nature of time factors which are not treated as unknown parameters to be estimated from the mortality data. Mortality projections are derived by means of the predictive distribution of the time index, i.e. its a posteriori distribution given past observations. This is the credibility feature of the proposed approach. New data feed this predictive distribution as they become available and so help to update mortality projections. This recognizes the dynamic aspect of mortality forecasting and avoids re-fitting the entire model based on new data. To the best of our knowledge, this dynamic updating approach has not been used so far and our numerical illustrations demonstrate its advantages compared to classical frequentist approaches.

The remainder of this paper is organized as follows. Section 2 gives a short introduction to evolutionary credibility models. Section 3 carefully presents the credibility model proposed to project future mortality. In Section 4, we discuss the covariance structure of the model and address the identifiability problem the model may encounter. Section 5 describes a two-step model fitting concept, which studies period and age effects separately. Section 6 is devoted to empirical illustrations. First, we fit the mortality experience of the G5 countries using our proposed methodology. Then, we study the index governing the payoff of the Swiss Re Kortis bond. Finally, we perform successive forecasts for the Belgian population to illustrate how newly available data can be

incorporated in revised forecasts. We compare the results to the official forecasts published yearly by the Federal Planning Bureau, the Belgian agency in charge of mortality projections.

## 2 Evolutionary credibility models

Following the book of Bühlmann and Gisler (2005) and focussing on the aspects that will be needed later on, this section gives a short introduction to evolutionary credibility modelling.

Consider a time series  $(r_t, \Theta_t)_{t \in \mathbb{N}}$  with a  $w$ -variate stochastic observation process  $(r_t)_{t \in \mathbb{N}}$  and a  $v$ -variate stochastic state (or risk profile) process  $(\Theta_t)_{t \in \mathbb{N}}$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . The state process  $(\Theta_t)$  is unobservable but shall follow a known dynamics. We are now at time  $T \in \mathbb{N}$  and the aim is to predict future states  $\Theta_{T+k}$ ,  $k \in \mathbb{N}$ , and conditional future expected observations  $\mathbb{E}[r_{T+k} \mid \Theta_{T+k}]$ ,  $k \in \mathbb{N}$ . The past observations  $r_1, \dots, r_T$  are the available information at time  $T$ .

Let all  $(r_t, \Theta_t)$ ,  $t \in \mathbb{N}$  be square integrable. Then the credibility estimator for  $\Theta_{T+k}$ , given the observations till time  $T$ , is defined as the orthogonal projection

$$\mu_{T+k|T} := \text{Pro}(\Theta_{T+k} \mid L(1, r_1, \dots, r_T)) \quad (2.1)$$

with respect to the set

$$L(1, r_1, \dots, r_T) := \left\{ a + \sum_{t=1}^T A_t r_t : a \in \mathbb{R}^v, A_t \in \mathbb{R}^{v,w} \right\}$$

in the Hilbert space of square integrable random variables. In other words,  $\mu_{T+k|T}$  is the unique element in the linear space  $L(1, r_1, \dots, r_T)$  that satisfies

$$\mathbb{E}[(\mu_{T+k|T} - \Theta_{T+k})(X - \Theta_{T+k})] = 0 \text{ for all } X \in L(1, r_1, \dots, r_T).$$

So  $\mu_{T+k|T}$  is the best linear predictor of  $\Theta_{T+k}$  in terms  $1, r_1, \dots, r_T$ .

Furthermore,  $(r_t, \Theta_t)$  is assumed to have a state-space representation of the form

$$r_t = G\Theta_t + W_t, \quad (2.2)$$

$$\Theta_{t+1} = F\Theta_t + V_t \quad (2.3)$$

with  $G \in \mathbb{R}^{v,v}$ ,  $F \in \mathbb{R}^{w,v}$  and white noise processes  $(W_t)$  and  $(V_t)$ . The two white noise processes shall be serially uncorrelated and also uncorrelated with each other. Their joint covariance matrix thus has the structure

$$\text{Cov} \left( \begin{pmatrix} V_t \\ W_t \end{pmatrix}, \begin{pmatrix} V_s \\ W_s \end{pmatrix} \right) = \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix}, \quad (2.4)$$

$Q \in \mathbb{R}^{v,v}$  and  $R \in \mathbb{R}^{w,w}$ , if  $t = s$  and zero else.

Under all these assumptions, the credibility estimator  $\mu_{T+k|T}$  for  $\Theta_{T+k}$  can be calculated in a recursive way, see Theorem 10.3 in Bühlmann and Gisler (2005). Starting

from an initial value  $\mu_{1|0} = \mathbb{E}[\Theta_1]$ , the estimate is sequentially updated by the newest observation through the recursive formula

$$\mu_{t|t} = \mu_{t|t-1} + A_t(r_t - G\mu_{t|t-1}) \quad (2.5)$$

for an appropriate matrix  $A_t$ . The step from  $t$  to  $t + 1$  then follows the evolution rule (2.3), i.e.

$$\mu_{t+1|t} = F\mu_{t|t}. \quad (2.6)$$

The credibility estimator  $\mu_{T+1|T}$  is obtained by iterating this procedure for  $t = 1, \dots, T$ . Finally,  $\mu_{T+k|T}$  and the credibility estimator for  $\mathbb{E}[r_{T+k} | \Theta_{T+k}] = G\Theta_{T+k}$  are given by

$$\begin{aligned} \mu_{T+k|T} &= F^{k-1}\mu_{T+1|T}, \\ \text{Pro}(G\Theta_{T+k} | L(1, r_1, \dots, r_T)) &= G\mu_{T+k|T}, \end{aligned}$$

respectively. This formula is also known as the Kalman recursion or the Kalman filter algorithm, cf. Brockwell and Davis (2006), and is implemented in the statistical software R.

A particular example of a stochastic process that can be expressed as an evolutionary credibility model is an autoregressive moving average (ARMA) process. Let  $(\Delta_t)$  follow an ARMA( $p, q$ ) process with AR parameters  $\phi_1, \dots, \phi_p$ , MA parameters  $\theta_1, \dots, \theta_q$  and innovation terms  $(Z_t)$ . Let  $d = \max\{p, q + 1\}$  and set  $\phi_k = 0$  for  $k > p$  and  $\theta_k = 0$  for  $k > q$ . According to Hamilton (1994),  $(\Delta_t)$  has the structure of (2.2) and (2.3):

$$\Delta_t = \underbrace{(1 \quad \theta_1 \quad \dots \quad \theta_{d-1})}_G \underbrace{\begin{pmatrix} H_t \\ H_{t-1} \\ \vdots \\ H_{t-d+1} \end{pmatrix}}_{\Theta_t} \quad (2.7)$$

with a  $d$ -dimensional state process  $\Theta_t = (H_t, H_{t-1}, \dots, H_{t-d+1})'$  following the state equation

$$\underbrace{\begin{pmatrix} H_t \\ H_{t-1} \\ \vdots \\ H_{t-d+1} \end{pmatrix}}_{\Theta_t} = \underbrace{\begin{pmatrix} \phi_1 & \dots & \phi_{d-1} & \phi_d \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix}}_F \underbrace{\begin{pmatrix} H_{t-1} \\ H_{t-2} \\ \vdots \\ H_{t-d} \end{pmatrix}}_{\Theta_{t-1}} + \underbrace{\begin{pmatrix} Z_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{V_t}. \quad (2.8)$$

### 3 A credibility model for mortality projection

On the basis of the concepts from the previous section, this section introduces and discusses mortality modelling within the evolutionary credibility framework. This allows us to apply credibility techniques and helps to make the forecasting more robust.

### 3.1 Age-specific improvement rates

We assume that we observe age-specific mortality statistics over an age range of  $x_1$  to  $x_n$ . So,  $n$  is the number of age groups included in the analysis. The mortality data relates to calendar years 1 to  $T$ , and we are now at the beginning of the year  $T + 1$ . For each age  $x = x_1, \dots, x_n$  and year  $t = 1, \dots, T$ , we calculate the crude death rate  $m_x(t)$  as the ratio of the number of deaths over the initial exposure-to-risk. Our aim is to project future rates  $m_x(T + 1), m_x(T + 2), \dots$  from the observed rates  $m_x(1), \dots, m_x(T)$ .

Mitchell et al. (2013) showed that it is advantageous to model the mortality rate changes rather than the mortality rate levels, because the dependence structure between ages of mortality is more accurately captured. Following the approach of Mitchell et al. (2013), we aim to model the log improvement rates

$$r_{xt} := \log m_x(t) - \log m_x(t - 1)$$

within the evolutionary modelling framework (2.2) and (2.3). Since Lee and Carter (1992) introduced their celebrated mortality model, it has become very common in the literature to assume that the general mortality trend and its effect on different age groups can be separated in a multiplicative way. So we assume that

$$r_{xt} = \beta_x \Delta_t + \epsilon_{xt}, \quad x = x_1, \dots, x_n, \quad (3.1)$$

where  $\Delta_t$  reflects the general development of mortality improvements with respect to calendar time, the parameters  $\beta_x \in \mathbb{R}$  measure the sensitivity of age group  $x$  with respect to the general mortality development, and  $\epsilon_{xt}$  is some noise. This multiplicative separation lets the parameters have very intuitive interpretations. Our model (3.1) looks similar to the Lee-Carter approach, yet recall that the  $r_{xt}$  describe the log mortality improvements and not the log mortality levels.

Another widespread approach in the mortality modelling literature is to interpret the general mortality trend  $\Delta_t$  as a stochastic time series. We propose to assume that  $\Delta_t$  is in the class of ARMA( $p, q$ ) processes of the form (2.7)-(2.8), i.e.

$$\Delta_t = (1 \quad \theta_1 \quad \dots \quad \theta_{d-1}) \Theta_t = G\Theta_t$$

and

$$\Theta_t = \begin{pmatrix} \phi_1 & \dots & \phi_{d-1} & \phi_d \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix} \Theta_{t-1} + \begin{pmatrix} Z_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} = F\Theta_{t-1} + V_t. \quad (3.2)$$

Apart from the fact that this is a rich and versatile class of time series processes, this way we obtain a model within the framework of evolutionary credibility models. Consequently,  $(r_{xt})$  has a state-space representation with an observation equation of the form

$$\begin{pmatrix} r_{x_1 t} \\ \vdots \\ r_{x_n t} \end{pmatrix} = \begin{pmatrix} \beta_{x_1} \\ \vdots \\ \beta_{x_n} \end{pmatrix} G\Theta_t + \begin{pmatrix} \epsilon_{x_1 t} \\ \vdots \\ \epsilon_{x_n t} \end{pmatrix}. \quad (3.3)$$



This representation shows that  $(r_{xt})$  still has the form of an evolutionary credibility model according to (2.2)-(2.3). Let the  $\Delta_t$  be Normally distributed with mean  $\delta$  and variance  $\sigma_\Delta^2$ ,

$$\Delta_t \sim \mathcal{N}(\delta, \sigma_\Delta^2).$$

We assume that the error terms  $\epsilon_{xt}$  have the distribution

$$\epsilon_{xt} \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

and are independent and also independent from the age-independent time trend  $\Delta = (\Delta_t)_{t \in \mathbb{N}}$ .

Conditional on  $\Delta_t$ , the mortality improvement rates  $(r_{xt})_{x=x_1}^{x_n}$  are independent for fixed  $t$ , but without conditioning on  $\Delta_t$ , they are serially correlated. Assuming that the noise is Normally distributed will allow us to obtain also predictive distributions for the credibility estimates, see Section 4.3.

Credibility estimations of future mortality rates depend on past observations through the recursive credibility formula and are sequentially updated as new observations enter the study. This recursive structure has positive implications from both the practical and the theoretical perspective. From the practical perspective, the updating of forecasts whenever a new observation arrives is numerically cheap. From the theoretical perspective, the recursive property means that there will be no extreme changes in the forecasts when a new data point is added. This means that the model will show some kind of robustness between successive forecasts.

As the specification (3.1) is not identifiable, some constraints are needed. In the remainder we always assume that

$$\sum_{x=x_1}^{x_n} \beta_x = 1. \tag{3.4}$$

Considering (3.1) and the assumptions made so far, we see that the correlation of mortality improvements at different ages,

$$\text{Corr}(r_{x_1t}, r_{x_2t}) = \frac{\beta_{x_1} \beta_{x_2} \sigma_\Delta^2}{\sqrt{\beta_{x_1}^2 \sigma_\Delta^2 + \sigma_\epsilon^2} \sqrt{\beta_{x_2}^2 \sigma_\Delta^2 + \sigma_\epsilon^2}},$$

may cover the entire range  $[-1, 1]$  when  $\sigma_\Delta^2$  and  $\sigma_\epsilon^2$  vary.

### 3.2 Aggregate mortality improvement rates

If we sum up the age-specific mortality improvement rates with respect to age, the coefficients  $\beta_x$  disappear because they add up to 1 according to (3.4). More precisely, we define the aggregate errors

$$\epsilon_{\bullet t} := \sum_{x=x_1}^{x_n} \epsilon_{xt},$$

which obey a Normal distribution with zero mean and variance

$$\sigma_{\bullet}^2 := n\sigma_{\epsilon}^2.$$

Aggregate errors are mutually independent and independent of  $\Delta_t$ . Considering (3.4), summing over  $x$  the identity (3.1) gives

$$r_{\bullet t} := \sum_{x=x_1}^{x_n} r_{xt} = \Delta_t + \epsilon_{\bullet t} \quad (3.5)$$

and it immediately follows that  $r_{\bullet t} \sim \mathcal{N}(\delta, \sigma_{\Delta}^2 + \sigma_{\bullet}^2)$ . In state-space form, regarding (3.2) and (3.3), (3.5) reads

$$r_{\bullet t} = G\Theta_t + \epsilon_{\bullet t} =: G\Theta_t + W_t \quad (3.6)$$

with state equation

$$\Theta_t = F\Theta_{t-1} + V_t. \quad (3.7)$$

The covariance matrix of the random vector  $(V'_t, W'_t)'$  is given by

$$\begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix},$$

where

$$Q := \begin{pmatrix} \sigma_Z^2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}, \quad (3.8)$$

$$R := \sigma_{\bullet}^2. \quad (3.9)$$

Here,  $\sigma_Z^2$  denotes the variance of the Gaussian white noise terms  $Z_t$ .

In a later stage of analysis, we propose a two-step model fitting procedure. We first consider the observed  $r_{\bullet 1}, \dots, r_{\bullet T}$  and we fit model (3.5). An analysis of (3.1) then follows. The advantage of this approach is that we are allowed to study the dynamics of  $\Delta_t$  describing improvement rates from the aggregate (3.5) involving the global improvement  $r_{\bullet t}$  and not the detailed age structure  $\beta_x$ .

### 3.3 Multi-population mortality improvement factors

The models (3.1) and (3.5) can be easily extended to integrate male and female improvement rates into a single model. This approach enforces consistency between genders, and in insurance applications it allows the actuary to evaluate potential diversification benefits between male and female future mortality improvements. More generally, several populations can be jointly modelled in a similar way. To fix the ideas we describe the

approach for a pair of populations, considering both genders in the same country index by  $i = m$  for males and  $i = f$  for females. In general,  $i = 1, 2 \dots$  indexes the populations to be considered.

Our model specification is as follows. Let  $r_{xt}^{(m)}$  denote the mortality improvement rates for males and let  $r_{xt}^{(f)}$  denote the corresponding mortality improvement rates for females from the same country. We now assume that model (3.1) applies to both genders, i.e. that

$$r_{xt}^{(i)} = \beta_x^{(i)} \Delta_t^{(i)} + \epsilon_{xt}^{(i)} \quad \text{for } i \in \{m, f\} \quad (3.10)$$

holds with a certain dependence structure between the two  $\Delta^{(i)}$ -processes specified later on. The parameters  $\beta_x^{(i)}$  add up to 1 for each  $i \in \{m, f\}$  in accordance with (3.4). Furthermore, the error terms  $\epsilon_{xt}^{(i)}$  in (3.10) are assumed to be mutually independent with distribution  $\mathcal{N}(0, \sigma_{i\epsilon}^2)$ . The corresponding aggregate structure is then given by

$$r_{\bullet t}^{(i)} = \Delta_t^{(i)} + \epsilon_{\bullet t}^{(i)} \quad \text{for } i \in \{m, f\}, \quad (3.11)$$

where  $\Delta_t^{(i)} \sim \mathcal{N}(\delta_i, \sigma_{i\Delta}^2)$ . Define the multivariate process

$$\mathbf{r}_{\bullet t} := \begin{pmatrix} r_{\bullet t}^{(m)} \\ r_{\bullet t}^{(f)} \end{pmatrix} = \begin{pmatrix} \Delta_t^{(m)} \\ \Delta_t^{(f)} \end{pmatrix} + \begin{pmatrix} \epsilon_{\bullet t}^{(m)} \\ \epsilon_{\bullet t}^{(f)} \end{pmatrix}.$$

Let

$$d := \max\{p_m, q_m + 1, p_f, q_f + 1\},$$

where  $(p_i, q_i)$  is the ARMA order of  $\Delta^{(i)}$ , and denote the gender specific ARMA parameters by an additional superscript  $(i)$ . Then, the state-space representations of  $\mathbf{r}_{\bullet t}$  is given by observation equation

$$\begin{aligned} \begin{pmatrix} r_{\bullet t}^{(m)} \\ r_{\bullet t}^{(f)} \end{pmatrix} &= \begin{pmatrix} 1 & \theta_1^{(m)} & \dots & \theta_{d-1}^{(m)} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & \theta_1^{(f)} & \dots & \theta_{d-1}^{(f)} \end{pmatrix} \begin{pmatrix} X_t^{(m)} \\ \vdots \\ X_{t-d+1}^{(m)} \\ X_t^{(f)} \\ \vdots \\ X_{t-d+1}^{(f)} \end{pmatrix} + \begin{pmatrix} \epsilon_{\bullet t}^{(m)} \\ \epsilon_{\bullet t}^{(f)} \end{pmatrix} \\ &=: \tilde{G}\tilde{\Theta}_t + \tilde{W}_t \end{aligned} \quad (3.12)$$

with state equation

$$\tilde{\Theta}_t = \begin{pmatrix} X_t^{(m)} \\ X_{t-1}^{(m)} \\ \vdots \\ X_{t-d+1}^{(m)} \\ X_t^{(f)} \\ X_{t-1}^{(f)} \\ \vdots \\ X_{t-d+1}^{(f)} \end{pmatrix} = \begin{pmatrix} \phi_1^{(m)} & \cdots & \phi_{d-1}^{(m)} & \phi_d^{(m)} & 0 & \cdots & 0 & 0 \\ 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & \phi_1^{(f)} & \cdots & \phi_{d-1}^{(f)} & \phi_d^{(f)} \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1}^{(m)} \\ X_{t-2}^{(m)} \\ \vdots \\ X_{t-d}^{(m)} \\ X_{t-1}^{(f)} \\ X_{t-2}^{(f)} \\ \vdots \\ X_{t-d}^{(f)} \end{pmatrix} + \begin{pmatrix} Z_t^{(m)} \\ 0 \\ \vdots \\ 0 \\ Z_t^{(f)} \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (3.13)$$

$=: \tilde{F}\tilde{\Theta}_{t-1} + \tilde{V}_t.$

Equations (3.12) and (3.13) combine the state-space representation (3.6) and (3.7) to a joint structure and dependency between the genders stems from the innovation errors  $Z_t^{(i)} \sim \mathcal{N}(0, \sigma_{iZ}^2)$ . They are correlated through a gender correlation parameter  $\gamma \in [-1, 1]$ , i.e.

$$\text{Cov}(Z_t^{(m)}, Z_t^{(f)}) = \gamma \sqrt{\sigma_{mZ}^2 \sigma_{fZ}^2}.$$

This defines the covariance matrices of  $(\tilde{V}_t', \tilde{W}_t')'$  as

$$Q := \begin{pmatrix} \sigma_{mZ}^2 & 0 & \cdots & 0 & \gamma \sqrt{\sigma_{mZ}^2 \sigma_{fZ}^2} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \gamma \sqrt{\sigma_{mZ}^2 \sigma_{fZ}^2} & 0 & \cdots & 0 & \sigma_{fZ}^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad (3.14)$$

$$R := \begin{pmatrix} \sigma_{m\bullet}^2 & 0 \\ 0 & \sigma_{f\bullet}^2 \end{pmatrix} \quad (3.15)$$

respectively.

The model allows for different degrees of homogeneity between the genders. A higher degree of homogeneity can be achieved by one of the following simplifying assumptions that reduce the number of model parameters.

- (S1)  $\Delta^{(m)}$  and  $\Delta^{(f)}$  are ARMA processes of the same order  $(p, q)$  and share the common ARMA parameters  $\phi_1, \dots, \phi_p$  and  $\theta_1, \dots, \theta_q$ .
- (S2) In addition to (S1), the variance parameters  $\sigma_{iZ}^2$  and  $\sigma_{i\bullet}^2$  do not depend on the gender  $i$ .

(S3) In addition to (S2), the drift parameter  $\delta_i$  does not depend on the gender  $i$ .

(S4) In addition to (S3), the gender correlation parameter is fixed at  $\gamma = 1$ .

All these assumptions can be easily incorporated into the above state-space representation. Activating all of them leads to the particular case  $\Delta_t^{(m)} = \Delta_t^{(f)}$  in which the mortality improvement factors applying to males and females are both functions of a single  $\Delta_t$ , i.e.

$$r_{xt}^{(i)} = \beta_x^{(i)} \Delta_t + \epsilon_{xt}^{(i)}, \quad i \in \{m, f\}. \quad (3.16)$$

This case is of particular interest as Carter and Lee (1992) suggested to use the same time index for both genders. To avoid long-run divergence in gender-specific mortality forecasts, Li and Lee (2005) further proposed to use the same  $\beta_x$  for all groups. Here, we nevertheless allow for gender-specific sensitivities  $\beta_x^{(i)}$  and leave the final decision to the user.

### 3.4 Mortality forecasting

Once the model parameters that enter the state-space representation are calibrated, the recursive formula according to Section 2 delivers credibility estimators for the future  $\Delta_{T+k}^{(i)}$ ,  $i \in \{m, f\}$ . Forecasts for the future death rates  $m_{x,T+k}^{(i)}$  and the corresponding one-year death probabilities  $q_{x,T+k}^{(i)}$  then directly follows. Precisely, by iterating the relationship

$$m_x^{(i)}(t) = m_x^{(i)}(t-1) \exp(\beta_x^{(i)} \Delta_t^{(i)} + \epsilon_{xt}^{(i)}),$$

we get

$$m_x^{(i)}(T+k) = m_x^{(i)}(T) \exp\left(\sum_{j=1}^k (\beta_x^{(i)} \Delta_{T+j}^{(i)} + \epsilon_{x,T+j}^{(i)})\right). \quad (3.17)$$

Inserting the credibility estimators  $\widehat{\Delta}_{T+j}^{(i)}$  gives point predictions

$$\widehat{m}_x^{(i)}(T+k) = m_x^{(i)}(T) \exp\left(\sum_{j=1}^k \widehat{\beta}_x^{(i)} \widehat{\Delta}_{T+j}^{(i)}\right) \quad (3.18)$$

of  $m_x^{(i)}(T+k)$ . Paths of future  $m_x^{(i)}(T+k)$  can be simulated by (3.17). The corresponding one-year death probabilities  $q_{x,T+k}^{(i)}$  and one-year survival probabilities  $p_{x,T+k}^{(i)}$  are easily obtained from

$$q_{x,T+k}^{(i)} = 1 - p_{x,T+k}^{(i)} = 1 - \exp(-m_x^{(i)}(T+k)).$$

Any quantity of interest can then be computed from these life tables.

## 4 Covariance structure and identifiability of the model

### 4.1 Single-population model

The ARMA process  $\Delta = (\Delta_t)_{t \in \mathbb{N}}$  is stationary Gaussian. The random vector  $(\Delta_1, \dots, \Delta_T)'$  thus obeys the multivariate Normal distribution with mean vector

$$\delta \mathbf{1}_T = (\delta, \dots, \delta)',$$

where  $\mathbf{1}_T = (1, \dots, 1)' \in \mathbb{R}^T$ , and covariance matrix of the Toeplitz form

$$\text{Cov}(\Delta_t, \Delta_s) = \rho_{|t-s|} \sigma_\Delta^2. \quad (4.1)$$

The  $\rho_h \in [-1, 1], h \in \mathbb{N}$ , with  $\rho_0 = 1$  are correlation parameters which are implicitly given through the underlying ARMA structure. In matrix notation (4.1) reads

$$\sigma_\Delta^2 \mathbf{C}_T = \sigma_\Delta^2 \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_1 \\ \rho_{T-1} & \dots & \rho_1 & 1 \end{pmatrix} \quad (4.2)$$

with correlation matrix  $\mathbf{C}_T \in \mathbb{R}^{T,T}$ . Note that the specification (4.1) has been also proposed by Sundt (1981) in a credibility context with an autoregressive structure, i.e. assuming

$$\rho_h = \rho^h, \quad h \in \mathbb{N}, \quad (4.3)$$

for some correlation parameter  $\rho \in [-1, 1]$ .

The model specification (3.5) directly implies that  $(r_{\bullet 1}, \dots, r_{\bullet T})'$  is multivariate Normal with mean vector

$$\delta \mathbf{1}_T = (\delta, \dots, \delta)' \quad (4.4)$$

and covariance matrix

$$\sigma_\bullet^2 I_T + \sigma_\Delta^2 \mathbf{C}_T. \quad (4.5)$$

In (4.5),  $I_T$  denotes the  $T \times T$  identity matrix. However, the covariance structure (4.5) may not be identifiable or in other words, not one-to-one with respect to its variance parameters. Let us make this point clear in the following example.

**Example 4.1.** *Assume that the time factor  $\Delta$  obeys the MA(1)-process*

$$\Delta_t = Z_t + \theta Z_{t-1}$$

*with independent innovation terms  $Z_t \sim \mathcal{N}(0, \sigma_Z^2)$ . Then, as*

$$\text{Var}(\Delta_t) = (1 + \theta^2) \sigma_Z^2 = \sigma_\Delta^2,$$

we see that  $\sigma_\Delta^2$  is implicitly given through the error variance  $\sigma_Z^2$  and the MA-parameter  $\theta$ . Moreover,

$$\mathbb{C}ov(\Delta_{t-1}, \Delta_t) = \theta \sigma_Z^2 = \frac{\theta}{1 + \theta^2} \sigma_\Delta^2$$

so that

$$\rho_h = \begin{cases} \frac{\theta}{1 + \theta^2} & \text{if } h = 1, \\ 0 & \text{if } h \geq 2. \end{cases}$$

If we replace  $\sigma_\Delta^2$  by  $\tilde{\sigma}_\Delta^2$ , the triples  $(\theta, \sigma_\Delta^2, \sigma_\bullet^2)$  and  $(\tilde{\theta}, \tilde{\sigma}_\Delta^2, \tilde{\sigma}_\bullet^2)$  with

$$\tilde{\sigma}_\bullet^2 = \sigma_\bullet^2 + \sigma_\Delta^2 - \tilde{\sigma}_\Delta^2$$

and  $\tilde{\theta}$  as the solution of

$$\frac{\tilde{\theta}}{1 + \tilde{\theta}^2} = \frac{\theta}{1 + \theta^2} \frac{\sigma_\Delta^2}{\tilde{\sigma}_\Delta^2}$$

will produce the same covariance matrix (4.5) provided that  $\tilde{\sigma}_\bullet^2 > 0$ . In fact, it is easy to verify that

$$\mathbb{C}ov(r_{\bullet,t-1}, r_{\bullet,t}) = \mathbb{C}ov(\Delta_{t-1}, \Delta_t) = \frac{\tilde{\theta}}{1 + \tilde{\theta}^2} \tilde{\sigma}_\Delta^2 = \frac{\theta}{1 + \theta^2} \sigma_\Delta^2$$

and

$$\mathbb{C}ov(r_{\bullet,t}, r_{\bullet,t}) = \tilde{\sigma}_\bullet^2 + \tilde{\sigma}_\Delta^2 = (\sigma_\bullet^2 + \sigma_\Delta^2 - \tilde{\sigma}_\Delta^2) + \tilde{\sigma}_\Delta^2 = \sigma_\bullet^2 + \sigma_\Delta^2.$$

All autoregressive processes, on the other hand, are identifiable but it is not clear whether these specifications appropriately explain the true dynamics of  $r_\bullet$ . The user should keep this point in mind when he decides to use the single-population model.

## 4.2 Multi-population model

As before, let us consider two populations to fix the ideas and assume that they consist in males and females in the same country. Identifiability is granted when considering  $(r_{\bullet,t}^{(m)}, r_{\bullet,t}^{(f)})$  and when some common structure is assumed.

**Lemma 4.2.** *Assume that (S1) holds, i.e. males and females share common ARMA parameters. Furthermore, suppose that the  $\Delta^{(i)}$  are causal. Then,*

$$\mathbb{C}ov(\Delta_t^{(m)}, \Delta_s^{(f)}) = \gamma \rho_{|t-s|} \sigma_\Delta^2, \quad (4.6)$$

where

$$\sigma_\Delta^2 = \sqrt{\sigma_{m\Delta}^2 \sigma_{f\Delta}^2}$$

and  $\rho_h$  is given by the gender individual covariance structure (4.1).

*Proof.* First assume that both  $\Delta^{(i)}$  follow a MA( $q$ ) process with parameters  $\theta_0, \dots, \theta_q$  with  $\theta_0 = 1$ . Then, for  $h \in \mathbb{N}$  and  $h \leq q$ , the covariance structure (3.14) of the innovation terms yields that

$$\begin{aligned} \text{Cov}(\Delta_{t+h}^{(i)}, \Delta_t^{(j)}) &= \text{Cov}\left(\sum_{k=0}^q \theta_k Z_{t+h-k}^{(i)}, \sum_{k=0}^q \theta_k Z_{t-k}^{(j)}\right) \\ &= \sum_{k=0}^q \sum_{l=0}^q \theta_k \theta_l \text{Cov}(Z_{t+h-k}^{(i)}, Z_{t-l}^{(j)}) \\ &= \sum_{k=0}^{q-h} \sum_{l=0}^q \theta_{k+h} \theta_l \text{Cov}(Z_{t-k}^{(i)}, Z_{t-l}^{(j)}) \\ &= \begin{cases} \left(\sum_{k=0}^{q-h} \theta_{k+h} \theta_k\right) \sigma_{iZ}^2, & i = j \\ \left(\sum_{k=0}^{q-h} \theta_{k+h} \theta_k\right) \gamma \sqrt{\sigma_{mZ}^2 \sigma_{fZ}^2}, & i \neq j. \end{cases} \end{aligned}$$

Taking  $h = 0$  gives

$$\sigma_{i\Delta}^2 = \text{Var}(\Delta_t^{(i)}) = \left(\sum_{k=0}^q \theta_k^2\right) \sigma_{iZ}^2$$

and therefore by  $\text{Cov}(\Delta_{t+h}^{(i)}, \Delta_t^{(i)}) = \rho_h \sigma_{i\Delta}^2$ ,

$$\rho_h = \left(\sum_{k=0}^{q-h} \theta_{k+h} \theta_k\right) \left(\sum_{k=0}^q \theta_k^2\right)^{-1}.$$

The case  $(-h) \in \mathbb{N}$  works in the same way such that (4.6) holds for any MA( $q$ ) process.

The result can be generalized using the moving average representation of ARMA( $p, q$ ) models. Since the  $\Delta^{(i)}$  are causal, there exist constants  $\psi_k \in \mathbb{R}$ ,  $k \in \mathbb{N}$ , with  $\sum_{k=0}^{\infty} |\psi_k| < \infty$  such that

$$\Delta_t^{(i)} - \delta_i = \sum_{k=0}^{\infty} \psi_k Z_{t-k}^{(i)} \quad (4.7)$$

almost surely and in  $\mathcal{L}^2$ . See Brockwell and Davis (2006) for details. Thus, (4.6) follows by taking  $q \rightarrow \infty$  and replacing  $\theta_k$  by  $\psi_k$  in the above calculations.  $\square$

Rewriting (4.6) in matrix notation, the covariance matrix between the vectors  $(\Delta_1^{(m)}, \dots, \Delta_T^{(m)})'$  and  $(\Delta_1^{(f)}, \dots, \Delta_T^{(f)})'$  is  $\gamma \sigma_{\Delta}^2 \mathbf{C}_T$ , where  $\mathbf{C}_T$  has been defined in (4.2). The structure directly follows from Lemma 4.2. Then, the gender-combined random vector

$$\mathbf{r}_{\bullet} = (r_{\bullet 1}^{(m)}, \dots, r_{\bullet T}^{(m)}, r_{\bullet 1}^{(f)}, \dots, r_{\bullet T}^{(f)})' \quad (4.8)$$

of past observed aggregate improvement rates is multivariate Normal with mean vector

$$\boldsymbol{\delta}_{\bullet} = (\delta_m, \dots, \delta_m, \delta_f, \dots, \delta_f)' \quad (4.9)$$



and covariance matrix

$$\boldsymbol{\Sigma}_\bullet = \begin{pmatrix} \sigma_{m\bullet}^2 I_T + \sigma_{m\Delta}^2 \mathbf{C}_T & \gamma \sigma_\Delta^2 \mathbf{C}_T \\ \gamma \sigma_\Delta^2 \mathbf{C}_T & \sigma_{f\bullet}^2 I_T + \sigma_{f\Delta}^2 \mathbf{C}_T \end{pmatrix}. \quad (4.10)$$

Compared to the single-population model, because of the additional parameter  $\gamma$ , there is no identifiability problem of  $\boldsymbol{\Sigma}_\bullet$  and the variance parameters anymore.

**Theorem 4.3.** *In addition to the conditions of Lemma 4.2, assume that the ARMA order  $(p, q)$  satisfies  $T - 1 > p + q$ . If  $\gamma \neq 0$ , then  $\boldsymbol{\Sigma}_\bullet$  is uniquely determined by the model parameters, i.e.  $\boldsymbol{\Sigma}_\bullet = \tilde{\boldsymbol{\Sigma}}_\bullet$  implies*

$$(\sigma_{m\bullet}^2, \sigma_{f\bullet}^2, \sigma_{m\Delta}^2, \sigma_{f\Delta}^2, \gamma, \rho_1, \dots, \rho_p) = (\tilde{\sigma}_{m\bullet}^2, \tilde{\sigma}_{f\bullet}^2, \tilde{\sigma}_{m\Delta}^2, \tilde{\sigma}_{f\Delta}^2, \tilde{\gamma}, \tilde{\rho}_1, \dots, \tilde{\rho}_p)$$

where  $\tilde{\boldsymbol{\Sigma}}_\bullet$  is the covariance matrix corresponding to the alternative parameters.

*Proof.* Assume that  $\boldsymbol{\Sigma}_\bullet = \tilde{\boldsymbol{\Sigma}}_\bullet$  and recall structure (4.2) of the correlation matrix  $\mathbf{C}_T$ . The structure of the diagonal block matrices of (4.10) implies that

$$\sigma_{i\Delta}^2 \rho_h = \tilde{\sigma}_{i\Delta}^2 \tilde{\rho}_h \quad (4.11)$$

for  $i \in \{m, f\}$  and  $h = 1, \dots, T - 1$ . The  $\rho_h$  and  $\tilde{\rho}_h$  are the entries of  $\mathbf{C}_T$  and  $\tilde{\mathbf{C}}_T$  respectively. Multiplying (4.11) for males and females and taking the square roots provide the identities

$$\sqrt{\sigma_{m\Delta}^2 \sigma_{f\Delta}^2 \rho_h} = \sqrt{\tilde{\sigma}_{m\Delta}^2 \tilde{\sigma}_{f\Delta}^2 \tilde{\rho}_h}, \quad (4.12)$$

$h = 1, \dots, T - 1$ . On the other hand, also the off diagonal matrices  $\gamma \sigma_\Delta^2 \mathbf{C}_T$  and  $\tilde{\gamma} \tilde{\sigma}_\Delta^2 \tilde{\mathbf{C}}_T$  agree. Therefore,

$$\gamma \sqrt{\sigma_{m\Delta}^2 \sigma_{f\Delta}^2 \rho_h} = \tilde{\gamma} \sqrt{\tilde{\sigma}_{m\Delta}^2 \tilde{\sigma}_{f\Delta}^2 \tilde{\rho}_h},$$

$h = 1, \dots, T - 1$ , and it follows from (4.12) and  $\gamma \neq 0 \neq \tilde{\gamma}$  that  $\gamma = \tilde{\gamma}$ . The condition  $T - 1 > p + q$  ensures that  $\rho_{h_1} \neq \rho_{h_2}$  and  $\tilde{\rho}_{h_1} \neq \tilde{\rho}_{h_2}$  for some  $h_1 \neq h_2$ . The identities of the further parameters follow step by step. Considering the diagonal entries of  $\gamma \sigma_\Delta^2 \mathbf{C}_T$  and its counterpart, we obtain

$$\gamma \sqrt{\sigma_{m\Delta}^2 \sigma_{f\Delta}^2} = \tilde{\gamma} \sqrt{\tilde{\sigma}_{m\Delta}^2 \tilde{\sigma}_{f\Delta}^2}$$

Since  $\gamma = \tilde{\gamma}$ ,

$$\sqrt{\sigma_{m\Delta}^2 \sigma_{f\Delta}^2} = \sqrt{\tilde{\sigma}_{m\Delta}^2 \tilde{\sigma}_{f\Delta}^2}$$

holds, which in turn proves

$$\rho_h = \tilde{\rho}_h, \quad h = 1, \dots, T - 1, \quad (4.13)$$

using (4.12). Relation (4.11) then provides

$$\sigma_{i\Delta}^2 = \tilde{\sigma}_{i\Delta}^2, \quad i \in \{m, f\}$$

and  $\sigma_{i\bullet}^2 = \tilde{\sigma}_{i\bullet}^2$  follows from

$$\sigma_{i\bullet}^2 + \sigma_{i\Delta}^2 = \tilde{\sigma}_{i\bullet}^2 + \tilde{\sigma}_{i\Delta}^2,$$

which are the diagonal entries of  $\sigma_{i\bullet}^2 I_T + \sigma_{i\Delta}^2 C_T$  and that of the alternative parameters respectively.  $\square$

### 4.3 Predictive Distribution

In this part, we assume that  $\Delta^{(m)}$  and  $\Delta^{(f)}$  are ARMA processes of the same order with gender-specific parameters  $\delta_i$ ,  $\sigma_{i\Delta}^2$  and  $\sigma_{i\bullet}^2$ . This corresponds to assumption (S1). The predictive distributions under the stronger assumptions (S2) to (S4) can be similarly obtained by making the parameters gender-common and/or setting  $\gamma$  to 1. For the single gender model, the appropriate submatrices must be chosen.

In applications, we are interested in the prediction of the future  $k$  years

$$\left( \Delta_{T+1}^{(m)}, \dots, \Delta_{T+k}^{(m)}, \Delta_{T+1}^{(f)}, \dots, \Delta_{T+k}^{(f)} \right)$$

given the past observed aggregate mortality improvement factors  $(r_{\bullet 1}^{(m)}, \dots, r_{\bullet T}^{(m)}, r_{\bullet 1}^{(f)}, \dots, r_{\bullet T}^{(f)})$ . We have already seen that credibility estimators, i.e. point predictions, follow from the recursive formula presented in Section 2. In the present setting of Normally distributed error terms, we can even specify the predictive distribution. This allows us to compute distribution related values, for instance confidence intervals of future mortality rates. To identify the predictive distribution we will need the following property of a Normal distribution.

**Lemma 4.4.** *Let  $(Y_1, Y_2) \sim \mathcal{N}((\mu_1, \mu_2), \Sigma)$  be a joint Normal distribution with*

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma'_{21} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix}.$$

*Then, conditional on  $Y_1 = y_1$ ,  $Y_2$  follows a Normal distribution with mean vector*

$$\mu_2 + \Sigma_{21} \Sigma_1^{-1} (y_1 - \mu_1) \tag{4.14}$$

*and covariance matrix*

$$\Sigma_2 - \Sigma_{21} \Sigma_1^{-1} \Sigma'_{21}. \tag{4.15}$$

The conditional expectation (4.14) indeed agrees with the recursive formula (2.5). Consider structure (2.2) and (2.3) with Normal white noise terms. Then for  $t = 1$ , we have  $\Theta_1 = V_1$  and  $r_1 = G\Theta_1 + W_1$ , i.e.

$$\Theta_1 \sim \mathcal{N}(\mathbb{E}[\Theta_1], Q) \quad \text{and} \quad r_1 \sim \mathcal{N}(G\mathbb{E}[\Theta_1], GQG' + R)$$

with  $\text{Cov}(\Theta_1, r_1) = QG'$ . Assigning  $r_1$  and  $\Theta_1$  the roles of  $Y_1$  and  $Y_2$  respectively, the conditional expectation of  $\Theta_1$  given  $r_1$  has the structure (4.14), which equals here

$$\mathbb{E}[\Theta_1] + QG' (GG' + R)^{-1} (r_1 - G\mathbb{E}[\Theta_1]).$$

This is exactly the recursive formula (2.5) for  $\mu_{1|1} = \text{Pro}(\Theta_1 | L(1, r_1))$  with initial value  $\mu_{1|0} = \mathbb{E}[\Theta_1]$ . The general structure  $\mu_{t|t}$  follows inductively.

In order to apply Lemma 4.4 to our mortality model, we require the distribution of the random vector

$$\left( r_{\bullet 1}^{(m)}, \dots, r_{\bullet T}^{(m)}, r_{\bullet 1}^{(f)}, \dots, r_{\bullet T}^{(f)}, \Delta_{T+1}^{(m)}, \dots, \Delta_{T+k}^{(m)}, \Delta_{T+1}^{(f)}, \dots, \Delta_{T+k}^{(f)} \right) \quad (4.16)$$

gathering past aggregate mortality improvement factors and future time indices. The  $T \times k$  correlation matrix  $\mathbf{C}_{T,k}$  of the past  $(\Delta_1^{(i)}, \dots, \Delta_T^{(i)})$  and the future  $(\Delta_{T+1}^{(i)}, \dots, \Delta_{T+k}^{(i)})$  up to horizon  $T+k$  is given by

$$\mathbf{C}_{T,k} = \begin{pmatrix} \rho_T & \rho_{T+1} & \dots & \rho_{T+k-1} \\ \rho_{T-1} & \rho_T & \dots & \rho_{T+k-2} \\ \vdots & \vdots & \dots & \vdots \\ \rho_1 & \rho_2 & \dots & \rho_k \end{pmatrix}.$$

The  $(t, l)$ -th entry in  $\mathbf{C}_{T,k}$  is  $\text{Corr}(\Delta_t^{(i)}, \Delta_{T+l}^{(i)}) = \rho_{T+l-t}$  and can be calculated by equation (5.2).

Further, define  $\mathbf{C}_{k,T} = \mathbf{C}'_{T,k}$ . The random vector (4.16) is multivariate Normal with mean vector  $\delta \mathbf{1}_{2T+2k}$  and covariance matrix

$$\begin{pmatrix} \sigma_{m\bullet}^2 I_T + \sigma_{m\Delta}^2 \mathbf{C}_T & \gamma \sigma_{\Delta}^2 \mathbf{C}_T & \sigma_{m\Delta}^2 \mathbf{C}_{T,k} & \gamma \sigma_{\Delta}^2 \mathbf{C}_{T,k} \\ \gamma \sigma_{\Delta}^2 \mathbf{C}_T & \sigma_{f\bullet}^2 I_T + \sigma_{f\Delta}^2 \mathbf{C}_T & \gamma \sigma_{\Delta}^2 \mathbf{C}_{T,k} & \sigma_{f\Delta}^2 \mathbf{C}_{T,k} \\ \sigma_{m\Delta}^2 \mathbf{C}_{k,T} & \gamma \sigma_{\Delta}^2 \mathbf{C}_{k,T} & \sigma_{m\Delta}^2 \mathbf{C}_k & \gamma \sigma_{\Delta}^2 \mathbf{C}_k \\ \gamma \sigma_{\Delta}^2 \mathbf{C}_{k,T} & \sigma_{f\Delta}^2 \mathbf{C}_{k,T} & \gamma \sigma_{\Delta}^2 \mathbf{C}_k & \sigma_{f\Delta}^2 \mathbf{C}_k \end{pmatrix}$$

and Lemma 4.4 can now be applied.

**Theorem 4.5.** *The predictive distribution for  $(\Delta_{T+1}^{(m)}, \dots, \Delta_{T+k}^{(m)}, \Delta_{T+1}^{(f)}, \dots, \Delta_{T+k}^{(f)})$  given  $(r_{\bullet 1}^{(m)}, \dots, r_{\bullet T}^{(m)}, r_{\bullet 1}^{(f)}, \dots, r_{\bullet T}^{(f)})$  is multivariate Normal with mean vector*

$$\begin{pmatrix} \delta_m \mathbf{1}_k \\ \delta_f \mathbf{1}_k \end{pmatrix} + \begin{pmatrix} \sigma_{m\Delta}^2 \mathbf{C}_{k,T} & \gamma \sigma_{\Delta}^2 \mathbf{C}_{k,T} \\ \gamma \sigma_{\Delta}^2 \mathbf{C}_{k,T} & \sigma_{f\Delta}^2 \mathbf{C}_{k,T} \end{pmatrix}^{-1} \begin{pmatrix} r_{\bullet}^{(m)} - \delta_m \mathbf{1}_T \\ r_{\bullet}^{(f)} - \delta_f \mathbf{1}_T \end{pmatrix} \quad (4.17)$$

and covariance matrix

$$\begin{pmatrix} \sigma_{m\Delta}^2 \mathbf{C}_k & \gamma \sigma_{\Delta}^2 \mathbf{C}_k \\ \gamma \sigma_{\Delta}^2 \mathbf{C}_k & \sigma_{f\Delta}^2 \mathbf{C}_k \end{pmatrix} - \begin{pmatrix} \sigma_{m\Delta}^2 \mathbf{C}_{k,T} & \gamma \sigma_{\Delta}^2 \mathbf{C}_{k,T} \\ \gamma \sigma_{\Delta}^2 \mathbf{C}_{k,T} & \sigma_{f\Delta}^2 \mathbf{C}_{k,T} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{m\bullet}^2 I_T + \sigma_{m\Delta}^2 \mathbf{C}_T & \gamma \sigma_{\Delta}^2 \mathbf{C}_T \\ \gamma \sigma_{\Delta}^2 \mathbf{C}_T & \sigma_{f\bullet}^2 I_T + \sigma_{f\Delta}^2 \mathbf{C}_T \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{m\Delta}^2 \mathbf{C}_{T,k} & \gamma \sigma_{\Delta}^2 \mathbf{C}_{T,k} \\ \gamma \sigma_{\Delta}^2 \mathbf{C}_{T,k} & \sigma_{f\Delta}^2 \mathbf{C}_{T,k} \end{pmatrix}. \quad (4.18)$$

In fact, (4.17) coincides with the credibility estimator. Since (4.17) equals

$$\mathbb{E} \left[ \left( \Delta_{T+1}^{(m)}, \dots, \Delta_{T+k}^{(m)}, \Delta_{T+1}^{(f)}, \dots, \Delta_{T+k}^{(f)} \right) \middle| r_{\bullet 1}^{(m)}, \dots, r_{\bullet T}^{(m)}, r_{\bullet 1}^{(f)}, \dots, r_{\bullet T}^{(f)} \right],$$

it is the Bayes estimator for the future  $\Delta$ -variables. Furthermore, it is also an affine function of  $\mathbf{r}_{\bullet}$ , i.e. a particular element of the Hilbert space  $L(1, \mathbf{r}_{\bullet 1}, \dots, \mathbf{r}_{\bullet T})$ . Thus, the claim follows from the definition of the credibility estimator (2.1).

## 5 Two-step model fitting concept

### 5.1 Model selection

As our models are fully specified, with Normally distributed components, the maximum likelihood approach is expected to deliver accurate estimations of the model parameters. We describe the fitting concept on the basis of the gender-combined model. As motivated in Section 3, we first fit the age-aggregate model to study period and age-effects separately. Let  $\mathbf{r}_{\bullet}$  gather the observed aggregate improvement rates as defined in (4.8). The log-likelihood function

$$\log L = -\frac{1}{2} \log |\boldsymbol{\Sigma}_{\bullet}| - \frac{1}{2} (\mathbf{r}_{\bullet} - \boldsymbol{\delta}_{\bullet}) \boldsymbol{\Sigma}_{\bullet}^{-1} (\mathbf{r}_{\bullet} - \boldsymbol{\delta}_{\bullet})' \quad (5.1)$$

of a  $2T$ -variate Normal distribution with mean vector  $\boldsymbol{\delta}_{\bullet}$  and covariance matrix  $\boldsymbol{\Sigma}_{\bullet}$  has to be maximized with respect to the model parameters. Notice that the input variables for the optimization are the ARMA parameters and the innovation variances  $\sigma_{iZ}^2$ , which appear in the state-space representation (3.12) to (3.15), rather than the correlation parameters  $\rho_h$  and the variance terms  $\sigma_{i\Delta}^2$ . The latter quantities can be deduced from the input variables using the identities

$$\rho_h = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+h}}{\sum_{j=0}^{\infty} \psi_j^2} \quad (5.2)$$

and

$$\sigma_{i\Delta}^2 = \sigma_{iZ}^2 \sum_{j=0}^{\infty} \psi_j^2,$$

where  $(\psi_j)_{j \in \mathbb{N}}$  comes from the corresponding MA( $\infty$ )-representation.

The model selection procedure in the gender-combined model then follows a backward approach. We start from a first model allowing for dynamics specific to each gender and we simplify it by activating the assumptions (S1) to (S4) step by step. In each step, ARMA( $p, q$ ) models of low orders are fitted to the underlying data. The candidate models are evaluated by using the Akaike information criterion with correction for finite samples (AICc), which is given by

$$\text{AICc} = -2 \log L + 2k + \frac{2k(k+1)}{T-k-1},$$

where  $k$  is the dimension of the parameter space. The AICc-best models proceed to the second stage of fitting. Nested models can be compared by means of likelihood-ratio tests. Simple models should be preferred if the difference of the log-likelihoods is not significant.

## 5.2 Age-specific structure

Given the parameters of the age-aggregate model, we calibrate the age-specific coefficients  $\beta_x^{(i)}$  and the residual variances  $\sigma_{i\epsilon}^2$  appearing in the age-specific model (3.10), which was

$$r_{xt}^{(i)} = \beta_x^{(i)} \Delta_t^{(i)} + \epsilon_{xt}^{(i)}.$$

Notice that for each age  $x$ , the random vector  $(r_{x1}^{(i)}, \dots, r_{xT}^{(i)})$  is multivariate Normal with mean vector

$$\beta_x^{(i)} \delta_i \mathbf{1}_T = (\beta_x^{(i)} \delta_i, \dots, \beta_x^{(i)} \delta_i)'$$

and covariance matrix

$$\sigma_{i\epsilon}^2 I_T + \beta_x^2 \sigma_{i\Delta}^2 \mathbf{C}_T.$$

The corresponding Normal log-likelihood function can thus be maximized with respect to the mean  $\beta_x^{(i)} \delta_i$  for each age  $x$  separately, which gives

$$\widehat{\beta_x^{(i)} \delta_i} = \frac{1}{T} \sum_{t=1}^T r_{xt}^{(i)}.$$

As the analysis of the aggregate mortality improvement rates  $r_{\bullet t}^{(m)}$  and  $r_{\bullet t}^{(f)}$  gives

$$\widehat{\delta}_i = \frac{1}{T} \sum_{t=1}^T r_{\bullet t}^{(i)},$$

we finally choose a plug-in estimator for  $\beta_x^{(i)} = \frac{\beta_x^{(i)} \delta_i}{\delta_i}$  as

$$\widehat{\beta}_x^{(i)} = \left( \sum_{t=1}^T r_{\bullet t}^{(i)} \right)^{-1} \sum_{t=1}^T r_{xt}^{(i)}, \quad (5.3)$$

which add up to 1. Hence, constraint (3.4) is satisfied. For  $\sigma_{i\epsilon}^2 = \frac{1}{n} \sigma_{i\bullet}^2$ , we choose

$$\widehat{\sigma}_{i\epsilon}^2 = \frac{1}{n} \widehat{\sigma}_{i\bullet}^2. \quad (5.4)$$

Country	(S1)	(S2)	(S3)	(S4)	optimal model
France	-	-	-	✓	AR(1)
Germany	-	-	✓	-	MA(3)
Japan	-	-	✓	-	ARMA(1,2)
UK	-	-	-	✓	MA(1)
USA	-	-	✓	-	AR(2)

Table 6.1: Results of the model selection procedure.

## 6 Empirical studies

### 6.1 Mortality projection models for the G5 countries

This section aims to analyze the pattern of mortality decline in the G5 countries (France, Germany, Japan, UK and US). We consider mortality data provided by the Human-Mortality-Database (2015). The analysis includes the ages  $x_1 = 21$  to  $x_n = 100$  ( $n = 80$ ) and the years from 1970 to the latest available time point. Observations before 1970 are not included since there is a structural break in the 70s as documented in Coelho and Nunes (2011). Thus,  $t = 1$  corresponds to the mortality improvement from calendar year 1970 to 1971 whereas  $T$  corresponds to the latest observation period. These choices are for illustrational purposes and our model is generally applicable for any choices of ages and years in fit.

Figure 6.1 displays the observed age-aggregated mortality improvements  $r_{\bullet t}^{(i)}$  where  $i = m$  for males and  $i = f$  for females of all G5 countries. Both series appear to be strongly correlated. Mortality statistics depicted in Figure 6.1 indicate negative correlation between  $r_{\bullet t}^{(i)}$  and  $r_{\bullet t+1}^{(i)}$ . This property is a consequence of the typical zigzag pattern, i.e. large improvements in mortality rates are followed by small improvements (or even declines) and vice versa. This apparent behavior also rules out time-invariant random effects in (3.1), i.e.  $\Delta_t^{(i)} = \Delta^{(i)}$  for all  $t \in \mathbb{N}$ , as this specification implies  $\text{Cov}(r_{\bullet t}^{(i)}, r_{\bullet s}^{(i)}) = \text{Var}(\Delta^{(i)}) > 0$  for all  $t \neq s$ . Hence,  $\Delta_t^{(i)} = \Delta^{(i)}$  would constrain  $r_{\bullet t}^{(i)}$  and  $r_{\bullet s}^{(i)}$  to be positively correlated among all years  $t$  and  $s$  which contradicts empirical evidence in Figure 6.1

### Results

The optimal models, i.e. the AICc-minimum models, for each country are summarized in Table 6.1 and Table 6.2. The checkmarks in Table 6.1 indicate which of the simplifying assumptions (S1) to (S4) is used in the final model. The corresponding model parameters are given in Table 6.2. We see that the flexibility of our model catches up the country-specific mortality characteristics. It is worth to mention that the simple white noise model, i.e. ARMA(0,0), is never optimal. This has great consequence for mortality forecasts as we point out next.

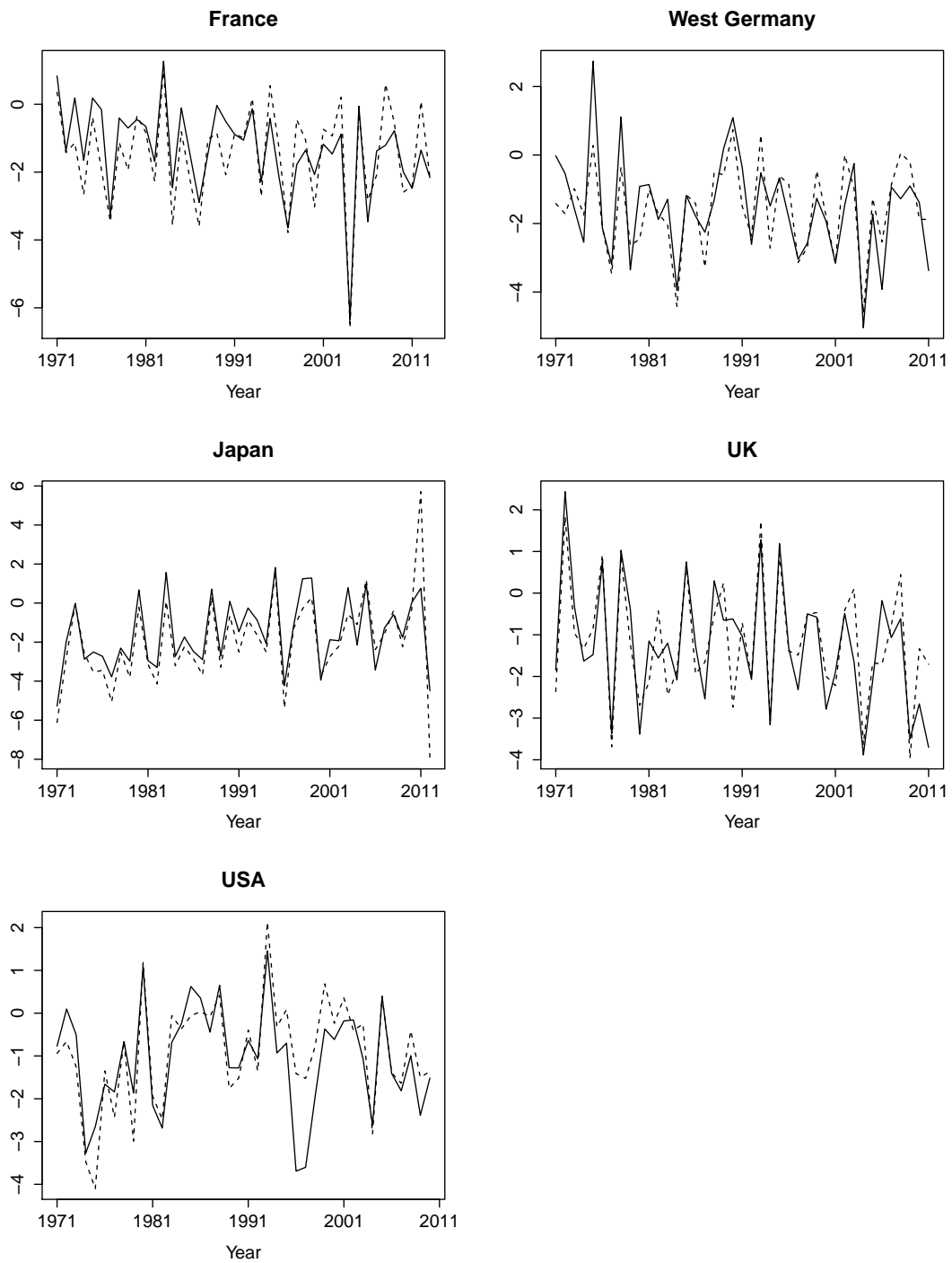


Figure 6.1: Age-aggregated mortality improvements  $r_{\bullet t}^{(i)}$  of males (solid line) and females (dashed line).

Country	$\delta_m$	$\delta_f$	$\sigma_{m\Delta}^2$	$\sigma_{f\Delta}^2$	$\sigma_{m\bullet}^2$	$\sigma_{f\bullet}^2$	$\gamma$	ARMA parameters
France	-1.410		1.636		0.328		1	$\phi_1 = -0.388$
Germany	-1.564		1.999		0.026		0.808	$\theta_1 = -0.290, \theta_2 = -0.089, \theta_3 = 0.646$
Japan	-1.770		5.113		0.004		0.862	$\phi_1 = 0.93, \theta_1 = -1.492, \theta_2 = 0.686$
UK	-1.212		1.703		0.348		1	$\theta_1 = -0.384$
USA	-0.996		1.899		0.015		0.882	$\phi_1 = -0.282, \phi_2 = 0.258$

Table 6.2: Model parameters.

## 6.2 Modelling Kortis Bond payoff

In December 2010, Swiss Re issued an index-linked bond through the Kortis Capital special purpose vehicle, paying quarterly floating coupons (at a margin of 5% above the three-month LIBOR value) and returning (part of) the principal at maturity. The principal is linked to an index measuring the divergence in mortality rates between male populations in England & Wales (representing about 90% of the UK population) and the US. Precisely, if the so-called Longevity Divergence Index Value (LDIV) is greater than 3.4%, then the principal of the bond is reduced linearly until full exhaustion of the principal if the LDIV is greater than 3.9%. The LDIV is computed as follows. First, an averaging period of eight years is used to calculate the observed improvement in death rates in each population for men at different ages, i.e.

$$1 - \left( \frac{m_x^{(i)}(t)}{m_x^{(i)}(t-8)} \right)^{\frac{1}{8}} = 1 - \exp \left( \frac{1}{8} \sum_{j=0}^7 r_{x,t-j}^{(i)} \right) \text{ where } i \in \{\text{US, UK}\}.$$

Second, an improvement index is calculated for each year and country over a specific age range using

$$I_t^{(\text{UK})} = \frac{1}{11} \sum_{x=75}^{85} \left( 1 - \exp \left( \frac{1}{8} \sum_{j=0}^7 r_{x,t-j}^{(\text{UK})} \right) \right)$$

and

$$I_t^{(\text{US})} = \frac{1}{11} \sum_{x=55}^{65} \left( 1 - \exp \left( \frac{1}{8} \sum_{j=0}^7 r_{x,t-j}^{(\text{US})} \right) \right).$$

Finally, the LDIV is calculated for year  $t$  as

$$LDIV_t = I_t^{(\text{UK})} - I_t^{(\text{US})}.$$

The principal of the Kortis bond is then reduced by

$$PRF = \max \left\{ \min \left\{ \frac{LDIV(2016) - 3.4\%}{3.9\% - 3.4\%}, 1 \right\}, 0 \right\},$$



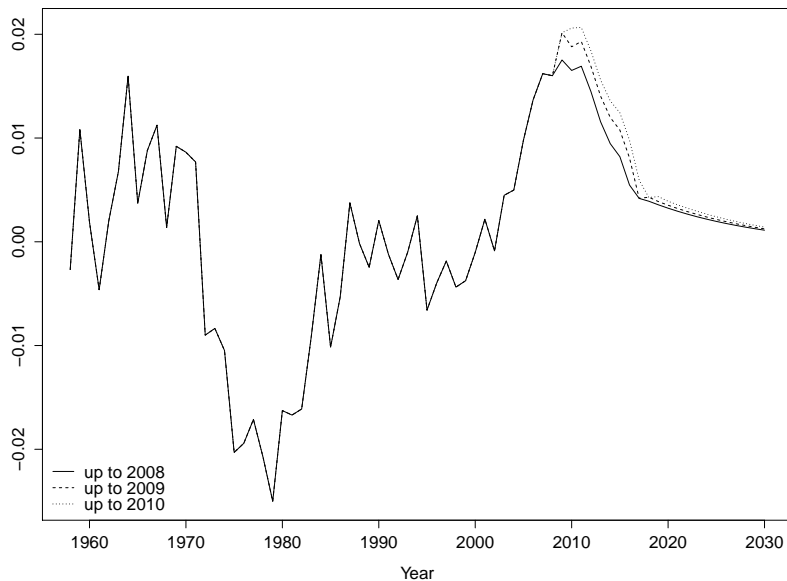


Figure 6.2: Historical path of the Kortis bond LDIV and forecast until 2030.

also known as principal reduction factor. For a detailed analysis of the Kortis bond, we refer the reader to Hunt and Blake (2015).

Let us now use the joint modelling of UK and US male population to calculate the performances of this bond. Specifically, the mortality projection model is used to forecast mortality for both populations and we calculate the LDIV in each case. The forecast is based on a joint US-UK model consisting in an ARMA(1,3) model with gender-specific mean and variance parameters and  $\gamma$  not fixed at 1, i.e. model assumption (S1) holds. Following Hunt and Blake (2015), the fit is based on observations of the years 1950 to 2008 and the ages 50 to 100. Then, forecasts are updated by including the observations for 2009 and 2010 respectively. The projection depicted in Figure 6.2 is very similar to the one displayed in Figure 8 of Hunt and Blake (2015). As these authors, we anticipate that the LDIV reaches a maximum near the end of the observation period and will decrease in value from that point. However, it is interesting to compare the impact of including new observations on this forecast. We see that the peak at the maximum becomes more marked when newly data are included in the analysis, without modifying the decreasing path after the maximum has been reached.

### 6.3 A detailed analysis for Belgium

In most countries, governmental agencies regularly publish mortality projections. In Belgium, the Federal Planning Bureau now produces projected life tables on an annual

basis, based on the most recent observations. However, standard forecasting approaches do not incorporate any smoothing procedure over time and this may cause some instability from one forecast to another. This is due to the fact that the model is entirely re-fitted based on an extended data set and that no connection is made between the successive projections. In this section, we show that the model proposed in this paper easily incorporates newly available data to revise current longevity forecasts.

We use Belgian data available from Statistics Belgium (<http://statbel.fgov.be/>), which also publishes official forecasts for the remaining life expectancy at the age of 65. To be considered is mortality after retirement age 65 for the calendar years 1970 to 2010. In a later stage of analysis we supplement these observations with calendar years 2011 and 2012 to study the robustness over successive forecasts.

Calibrating the gender-combined model leads to the model where all simplifying assumptions (S1) to (S4) are activated. In the optimal model,  $\Delta^{(m)}$  and  $\Delta^{(f)}$  are ARMA(1,1) processes with parameters  $\phi_1 = -0.999$  and  $\theta_1 = 0.429$ . Further model parameters are shown in Table 6.3.

Table 6.3 also demonstrates estimates for the age-aggregate mortality improvement model implicitly given by the Lee-Carter model. Recall that in the Lee-Carter framework, the log death rates  $\log m_x^{(i)}(t)$ ,  $i \in \{m, f\}$ , are decomposed by a principal component analysis into  $\alpha_x^{(i)} + \beta_x^{(i)} \kappa_t^{(i)}$  where the time factor  $\kappa_t^{(i)}$  obeys an ARIMA dynamics. Therefore,

$$\sum_{x=x_1}^{x_n} (\log m_x^{(i)}(t) - \log m_x^{(i)}(t-1)) = \sum_{x=x_1}^{x_n} \beta_x^{(i)} (\kappa_t^{(i)} - \kappa_{t-1}^{(i)}) = \kappa_t^{(i)} - \kappa_{t-1}^{(i)}$$

for  $i \in \{m, f\}$ . As in the majority of empirical studies conducted with the Lee-Carter model, we assume that  $\kappa_t^{(i)}$  obeys the random walk with drift model

$$\kappa_t^{(i)} - \kappa_{t-1}^{(i)} = \delta_{i\kappa} + S_t^{(i)}$$

with independent error components  $S_t^{(i)} \sim \mathcal{N}(0, \sigma_{i\kappa}^2)$ . Furthermore, the residual variance between the observed and fitted model is denoted by  $\sigma_{i\circ}^2$  which is the analog term to  $\sigma_{i\bullet}^2$  in our model. Even though both models are based on a similar structure, the differences in the estimated values are remarkable. Drift parameters clearly vary from those of the Lee-Carter model. What is even more important is how the total variances  $\sigma_{i\bullet}^2 + \sigma_{i\Delta}^2$  and  $\sigma_{i\circ}^2 + \sigma_{i\kappa}^2$  are allocated in the two models. While the Lee-Carter mortality improvement model gives more weight to the measurement variance  $\sigma_{i\circ}^2$ , the innovation variance  $\sigma_{i\Delta}^2$  is dominating in our model. Notice that the innovation error affects all ages through the sensitivity factor  $\beta_x^{(i)}$ .

## 6.4 Forecasts of period life expectancies

We illustrate the forecasts of our mortality model on the basis of period life expectancies. Given the predicted survival probabilities  $\hat{p}_{x,T+k}$ , see Subsection 3.4, the period life

	$\delta_m$	$\delta_f$	$\sigma_{m\Delta}^2$	$\sigma_{f\Delta}^2$	$\sigma_{m\bullet}^2$	$\sigma_{f\bullet}^2$
Our model	-0.554		1.920		0.180	
Lee-Carter	-0.434	-0.522	0.330	0.664	1.790	0.699

Table 6.3: Estimated mean and variance parameters. For the Lee-Carter models, the values are the estimates for  $\delta_{i\kappa}$ ,  $\sigma_{i\kappa}^2$  and  $\sigma_{i\circ}^2$  respectively.

	Our model	Lee-Carter with $m_x(2010)$	Lee-Carter with $\alpha_x$
male	22.5540	21.9329	19.5003
female	25.3981	25.3079	23.2666

Table 6.4: Point forecasts of period life expectancy in 2050

expectancy  $\hat{e}_{65}(T+k)$  at age 65 in calendar year  $T+k$  can be calculated using the formula

$$\hat{e}_{65}(T+k) = \frac{1}{2} + \sum_{j \geq 1} \prod_{l=0}^{j-1} \hat{p}_{65+l, T+k}. \quad (6.1)$$

The predicted mortality improvements are applied on the last observation  $m_x^{(i)}(2010)$ . Table 6.4 shows point predictions  $\hat{e}_{65}(2050)$  and the same point forecasts obtained by the Lee-Carter model. Considering the Lee-Carter forecast, applying the mortality reduction factors to the last observations  $m_x(2010)$  greatly affects the projected  $e_{65}(T+k)$ . In the remainder, Lee-Carter forecasts use  $m_x^{(i)}(T)$  as an initial value instead of the offsets  $\alpha_x^{(i)}$ . In this case, the forecasts roughly agree.

To conclude, let us show that the model proposed in the present thesis solves the robustness issue mentioned in the introduction, when applied sequentially over the years. To this end, we fit the model using data up to 2010 and update the predictive distribution by using data up to years 2011 and 2012. This provides three forecasts of future mortality that we compare together as well as to the Lee-Carter forecasts and the three official forecasts published by Statistics Belgium over the same period. Figure 6.3 shows the forecasts for  $e_{65}(T+1), \dots, e_{65}(2015)$  with  $T = 2010, 2011$  and  $2012$ , starting from the latest available  $e_{65}(T)$ . It can be clearly seen that differences in the initial values are stabilized over time for our model, whereas forecasts by Lee-Carter are just straight lines starting from the different initial values. The robustness is a consequence of the underlying ARMA structure, i.e. mortality improvements not being independent in time. More precisely, the estimated autocorrelation function of the time index is negative for lags of size one. Thus,  $\Delta_t^{(i)}$  is likely to be followed by  $\Delta_{t+1}^{(i)}$  going into the opposite direction. By (3.17), the deviations cancel out. On the other hand, mortality improvements

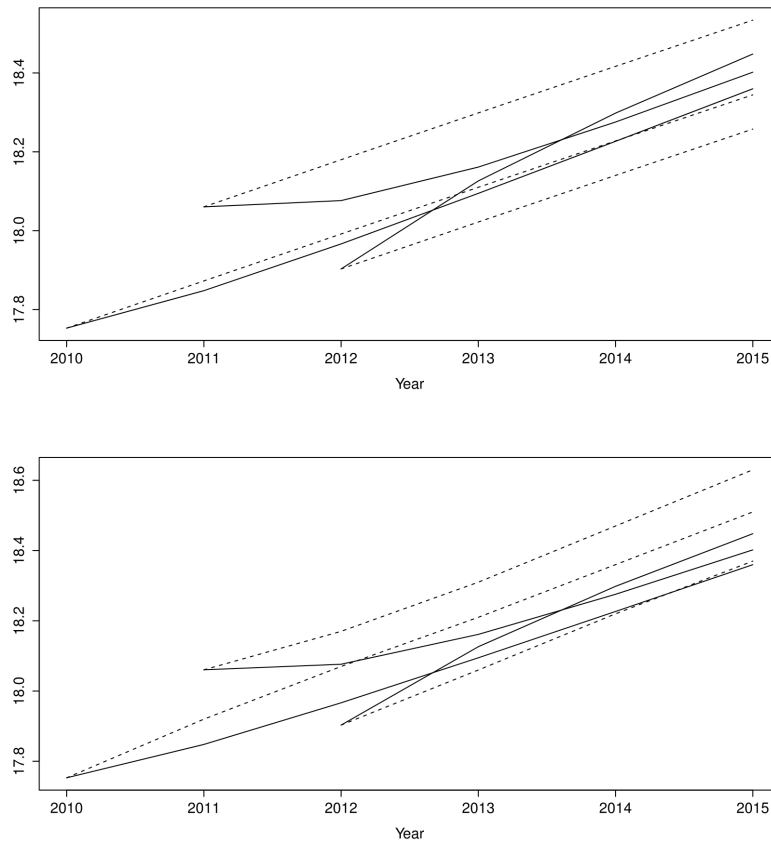


Figure 6.3: Comparison of predicted paths of period life expectancies. In the top figure we compare predictions of our model (solid line) to Lee-Carter (dashed line). In the bottom figure, our predictions are compared to the predictions of Statistics Belgium (dashed line). The three consecutive lines are based on data up to 2010, 2011 and 2012 respectively.

are independent under the Lee-Carter model or under any random walk model. Thus, outliers remain and strongly impact the future life expectancy.

## 7 Conclusion

We showed how evolutionary credibility modeling can be successfully applied for the forecasting of mortality in single and multiple populations. The focus on mortality improvements rather than mortality levels combined with the recursive ARMA structure that we used leads to robustness of the forecasts for successive years. Our empirical analysis of Belgian data confirms this advantage compared to models of Lee-Carter type. With the Kortis bond we gave another example where such robustness is a desirable

property. By using Normal distribution assumptions, the credibility estimator exactly equals the Bayes estimator in our model. Relaxing the distribution assumptions would require sophisticated estimation techniques for the Bayes estimator, while the credibility estimator keeps its nice and simple form. Therefore we think that credibility theory is a very useful technique for mortality forecasting of multiple populations, far beyond the scope of this paper.

## Acknowledgements

Edo Schinzinger thanks for the financial support granted by the “DFG Research Training Group 1100”.

Michel Denuit gratefully acknowledges the financial support from the UNIL “Chaire Pensions et Longévité” financed by Retraites Populaires, directed by Professor François Dufresne.

## References

- Aleksic, M.-C. and M. Börger (2011). Coherent projections of age, period, and cohort dependent mortality improvements. Technical report, Discussion Paper, University of Ulm.
- Brockwell, P. J. and R. A. Davis (2006). *Time series: theory and methods*. Springer.
- Bühlmann, H. and A. Gisler (2005). *A course in credibility theory and its applications*. Springer.
- Carter, L. R. and R. D. Lee (1992). Modeling and forecasting us sex differentials in mortality. *International Journal of Forecasting* 8(3), 393–411.
- Coelho, E. and L. C. Nunes (2011). Forecasting mortality in the event of a structural change. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 174(3), 713–736.
- Czado, C., A. Delwarde, and M. Denuit (2005). Bayesian poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics* 36(3), 260–284.
- Debón, A., F. Montes, and F. Martínez-Ruiz (2011). Statistical methods to compare mortality for a group with non-divergent populations: an application to spanish regions. *European Actuarial Journal* 1(2), 291–308.
- Delwarde, A., M. Denuit, M. Guillen, A. Vidiella, et al. (2006). Application of the poisson log-bilinear projection model to the g5 mortality experience. *Belgian Actuarial Bulletin* 6(1), 54–68.
- Giroso, F. and G. King (2008). *Demographic forecasting*. Princeton University Press.

- Hamilton, J. D. (1994). *Time series analysis*, Volume 2. Princeton university press Princeton.
- Human-Mortality-Database (2015). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at [www.mortality.org](http://www.mortality.org) or [www.humanmortality.de](http://www.humanmortality.de) (data downloaded on May 2nd, 2015).
- Hunt, A. and D. Blake (2014). A general procedure for constructing mortality models. *North American Actuarial Journal* 18(1), 116–138.
- Hunt, A. and D. Blake (2015). Modelling longevity bonds: Analysing the swiss re kortis bond. *Insurance: Mathematics and Economics*, <http://dx.doi.org/10.1016/j.insmatheco.2015.03.017>.
- Kogure, A., K. Kitsukawa, and Y. Kurachi (2009). A bayesian comparison of models for changing mortalities toward evaluating longevity risk in japan. *Asia-Pacific Journal of Risk and Insurance* 3(2), 1–22.
- Kogure, A. and Y. Kurachi (2010). A bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. *Insurance: Mathematics and Economics* 46(1), 162–172.
- Lee, R. D. and L. R. Carter (1992). Modeling and forecasting us mortality. *Journal of the American statistical association* 87(419), 659–671.
- Li, J. (2013). A poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population studies* 67(1), 111–126.
- Li, J. (2014). An application of MCMC simulation in mortality projection for populations with limited data. *Demographic Research* 30(1), 1–48.
- Li, N. and R. Lee (2005). Coherent mortality forecasts for a group of populations: An extension of the lee-carter method. *Demography* 42(3), 575–594.
- Mitchell, D., P. Brockett, R. Mendoza-Arriaga, and K. Muthuraman (2013). Modeling and forecasting mortality rates. *Insurance: Mathematics and Economics* 52(2), 275–285.
- Pedroza, C. (2006). A bayesian forecasting model: Predicting us male mortality. *Biostatistics* 7(4), 530–550.
- Russolillo, M., G. Giordano, and S. Haberman (2011). Extending the lee-carter model: a three-way decomposition. *Scandinavian Actuarial Journal* 2011(2), 96–117.
- Sundt, B. (1981). Recursive credibility estimation. *Scandinavian Actuarial Journal* 1981(1), 3–21.

Yang, S. S. and C.-W. Wang (2013). Pricing and securitization of multi-country longevity risk with mortality dependence. *Insurance: Mathematics and Economics* 52(2), 157–169.

Zhou, R., J. S.-H. Li, and K. S. Tan (2013). Pricing standardized mortality securitizations: A two-population model with transitory jump effects. *Journal of Risk and Insurance* 80(3), 733–774.