I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
( I S B A )

UNIVERSITÉ CATHOLIQUE DE LOUVAIN


D I S C U S S I O N
P A P E R

2014/21

# An Evolutionary Credibility Model of Lee-Carter Type for Mortality Improvement Rates

SCHINZINGER, E., DENUIT, M. and M. CHRISTIANSEN

# AN EVOLUTIONARY CREDIBILITY MODEL OF LEE-CARTER TYPE FOR MORTALITY IMPROVEMENT RATES

Edo Schinzinger
Research Training Group 1100
University of Ulm
Ulm, Germany

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

Marcus C. Christiansen
Institute of Insurance Science
University of Ulm
Ulm, Germany

May 28, 2014

**Abstract**

If fitted each year based on the latest available information, the forecasts produced by standard mortality projection models are likely to be unstable. To avoid this problem, the present paper proposes an evolutionary credibility model inspired by Lee and Carter (1992). Instead of modeling the mortality rate levels, the time series of mortality rate changes, or mortality improvement rates are considered and expressed in terms of a single mortality index, 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 the underlying time index. Empiricial illustrations based on Belgian mortality statistics demonstrate the practical interest of the approach proposed in the present paper.

*Key words and phrases:* Life insurance; mortality projection; Lee-Carter model; predictive distribution.

# 1 Introduction and motivation

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. In most countries, governmental agencies regularly publish mortality projections. In Belgium for instance, 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.

Considering the elegant approach to mortality forecasting pioneered by Lee and Carter (1992), an important aspect of the methodology is that the time factor is intrinsically viewed as a stochastic process and Box-Jenkins techniques are used to estimate and forecast it within an ARIMA time series model. From this forecast of the general level of mortality, the actual age-specific death rates are derived using the estimated age effects. This in turn yields projected life expectancies.

Precisely, starting from a matrix of centered death rates (on the log scale) indexed by calendar time (column) and age (row), the time index can be obtained by singular value decomposition (SVD) as the eigenvector associated with the largest eigenvalue. As time passes, new columns are added to this matrix which impacts on the estimated time index. Regression models can also be used instead of SVD but the issue remains: increasing the observation period modifies the time index extracted from the mortality data. In a second step, this time index is considered as the realization of a time series and ARIMA models are used to extrapolate it to the future. In the majority of empirical studies, the simple ARIMA(0,1,0), or random walk with drift model is used. However, the estimation of the drift in the classical random walk model for the time index only depends on the value of the time index for the first and the last calendar years. Therefore, including a new observation with a very different time index value greatly impacts on the estimated drift, which in turn results in very different mortality projections.

Most often, the modeling still proceeds in two steps: first the time index is estimated as if it was a real-valued parameter, and then it is extrapolated using Box-Jenkins methodology. The random nature of the unobservable time index is, thus, disregarded, which may bias the analysis. As possible incoherence may arise from this two-step procedure, Czado et al. (2005) integrated both steps into a Bayesian model 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, 2010) and Li (2014) for related works. However, 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 stability over time while keeping the computational issues relatively easy and allowing for the flexibility of ARIMA modeling. It is worth stressing that the time index is here treated as such, and not as a parameter to be estimated from past mortality statistics using SVD or regression techniques before entering time series models. In this way, we recognize the hidden nature of the time index and its intrinsic randomness.

Whereas most mortality studies consider both genders separately, the model proposed in this paper combines male and female mortality statistics. While ensuring model identification, this is particularly useful in practical applications as insurance portfolios mix both genders and the corresponding time indices are strongly correlated. Separate analyses could then lead to miss this strong dependence pattern, which considerably reduces possible diversification effects inside the portfolio.

The remainder of this paper is organized as follows. Section 2 carefully presents the credibility model proposed to project future mortality. Section 3 is devoted to numerical illustrations based on Belgian data. Section 4 describes mortality forecasting and discusses the numerical results.

# 2 A credibility model for mortality projection

## 2.1 Age-specific improvement rates

### 2.1.1 Available data

Assume that we have at our disposal age-specific mortality statistics over an age range $\{x_1, x_1+1, \ldots, x_n\}$. Here, $n$ is the number of ages included in the analysis. These data relate to calendar years 1 to $T$ and we are now at the end of year $T$, i.e. at time $T+1$. For each age $x \in \{x_1, x_1+1, \ldots, x_n\}$ and year $t \in \{1, \ldots, T\}$, we have the value of the crude (central) death rate $m_x(t)$ obtained as the ratio of the number of deaths over the initial exposure-to-risk. Our aim is to project the past $m_x(t)$ to get future $m_x(T+k)$, $k = 1, 2, \ldots$, entering actuarial calculations.

### 2.1.2 Log-bilinear decomposition for death rates

After Lee and Carter (1992), the central death rate $m_x(t)$ is decomposed in a log-bilinear way on the log-scale into

$$\ln m_x(t) = \alpha_x + \beta_x \kappa_t. \tag{2.1}$$

In (2.1), parameters $\alpha_x$ describe the general shape of the mortality curve, the time index $\kappa_t$ reflects the general level of mortality and the age-specific component $\beta_x$ represents how rapidly or slowly mortality at each age varies when the general level of mortality changes. As the specification (2.1) is not identifiable, some constraints are needed. This is why we assume in the remainder of the paper that

$$\sum_{x=x_1}^{x_n} \beta_x = 1. \tag{2.2}$$

### 2.1.3 From Lee-Carter to dynamic credibility modeling

Define

$$r_{xt} = \ln \frac{m_x(t)}{m_x(t-1)}$$

as the log improvement rate in mortality at age $x$ from year $t-1$ to year $t$. Recently, several authors suggested to target $r_{xt}$ to forecast future mortality, instead of the death rates. See, e.g., Mitchell et al. (2013) or Börger and Aleksic (2013). Now, under (2.1), we get

$$r_{xt} = \ln m_x(t) - \ln m_x(t-1) = \beta_x(\kappa_t - \kappa_{t-1}). \tag{2.3}$$

This leads us proposing the following evolutionary credibility model that is based on log mortality rate changes rather than levels:

$$r_{xt} = \beta_x \Delta_t + \epsilon_{xt} \tag{2.4}$$

where $\Delta_t \sim \mathcal{N}or(\delta, \sigma_\Delta^2)$ plays the role of the first difference of the time index $\kappa_t$. Model (2.4) appears to be a natural credibility counterpart to (2.3): given $\Delta_t$, the mortality improvement rates $\{r_{xt},\ x = x_1, \ldots, x_n\}$ are independent but become unconditionally serially correlated. Past observations of $r_{xt}$ modify the distribution of future $\Delta_t$, impacting future mortality.

In line with the Lee-Carter methodology, we assume that $\Delta_t$ appearing in (2.4) obeys some ARMA dynamics and the error terms $\epsilon_{xt} \sim \mathcal{N}or(0, \sigma_\epsilon^2)$ are independent and independent from $\Delta_t$. The parameters $\beta_x$ still measure the sensitivity of mortality at age $x$ with respect to calendar time and are assumed to fulfill (2.2).

## 2.2 Aggregate mortality improvement rates

Considering (2.2), summing over $x$ the identity (2.4) gives

$$r_{\bullet t} = \sum_{x=x_1}^{x_n} r_{xt} = \Delta_t + \epsilon_{\bullet t} \sim \mathcal{N}or(\delta, \sigma_\Delta^2 + \sigma_\bullet^2), \tag{2.5}$$

where the aggregate errors

$$\sum_{x=x_1}^{x_n} \epsilon_{xt} = \epsilon_{\bullet t} \sim \mathcal{N}or(0, \sigma_\bullet^2) \text{ with } \sigma_\bullet^2 = n\sigma_\epsilon^2$$

are mutually independent and independent of $\Delta_t$.

In the aggregate approach, we first consider the observed $r_{\bullet 1}, \ldots, r_{\bullet T}$ and we fit model (2.5). The advantage of this approach is that we are allowed to study the dynamics of $\Delta_t$ describing improvement rates from the aggregate (2.5) involving the global improvement $r_{\bullet t}$ and not the detailed age structure $\beta_x$.

## 2.3 Covariance structure

The state process $\{\Delta_t,\ t = 1, 2, \ldots\}$ is assumed to be stationary Gaussian. The random vector $(\Delta_1, \ldots, \Delta_T)'$ thus obeys the multivariate Normal distribution with mean vector

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

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

$$\mathbb{C}\text{ov}[\Delta_t, \Delta_s] = \rho_{|t-s|}\sigma_\Delta^2 \tag{2.6}$$

for correlation parameters $\rho_h \in [-1,1]$, $h = 1,2,\ldots$, and $\rho_0 = 1$. In matrix notation (2.6) reads

$$\sigma_\Delta^2 \boldsymbol{C}_T = \sigma_\Delta^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1 \end{pmatrix} \tag{2.7}$$

with correlation matrix $\boldsymbol{C}_T$. Note that specification (2.6) has been also proposed by Sundt (1981) in a credibility context with an autoregressive structure, i.e. assuming

$$\rho_h = \rho^h, \;\; h = 1,2,\ldots \tag{2.8}$$

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

The model specification (2.5) directly implies that $(r_{\bullet 1},\ldots,r_{\bullet T})'$ is multivariate Normal with mean vector

$$\delta \mathbf{1}_T = (\delta,\ldots,\delta)' \tag{2.9}$$

and covariance matrix

$$\sigma_\bullet^2 \boldsymbol{I}_T + \sigma_\Delta^2 \boldsymbol{C}_T, \tag{2.10}$$

respectively. In (2.10), $\boldsymbol{I}_T$ denotes the $T \times T$ identity matrix.

However, the covariance structure (2.10) may not be identifiable or in other words, not one-to-one with respect to its parameters. For instance, replace $\sigma_\Delta^2 > 0$ with $\tilde{\sigma}_\Delta^2 = 2\sigma_\Delta^2$. Then, moving to $\tilde{\rho}_h = \frac{1}{2}\rho_h$, $h \geq 1$, and $\tilde{\sigma}_\bullet^2 = \sigma_\bullet^2 - \sigma_\Delta^2$ leads to the same covariance matrix as for $(\sigma_\Delta^2, \sigma_\bullet^2, \rho_h)$. Although special choices of $\rho_h$ are admissible, such as the autoregressive structure (2.8), it is not clear whether these specifications appropriately explain the true dynamics of $(r_{\bullet 1},\ldots,r_{\bullet T})'$. In order to solve this issue, we now consider both genders together, as explained next.

## 2.4 Gender-combined mortality improvement factors

Instead of considering males and females separately, we now integrate both gender-specific improvement rates into a single model. In addition to ensuring identifiability of the covariance structure, this approach also allows to evaluate potential diversification benefits between male and female future mortality improvements.

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 the model (2.4) applies to both genders, i.e.

$$r_{xt}^{(i)} = \beta_x^{(i)} \Delta_t^{(i)} + \epsilon_{xt}^{(i)} \quad \text{for } i \in \{m,f\} \tag{2.11}$$

where the parameters $\beta_x^{(i)}$ sum to 1 for each $i \in \{m,f\}$ in accordance with (2.2). The error terms $\epsilon_{xt}^{(i)}$ in (2.11) are assumed to be mutually independent and to follow the $\mathcal{N}or(0,\sigma_{i\epsilon}^2)$ distribution. The corresponding aggregate structure is then given by

$$r_{\bullet t}^{(i)} = \Delta_t^{(i)} + \epsilon_{\bullet t}^{(i)} \quad \text{for } i \in \{m,f\}. \tag{2.12}$$

The dynamics for the state processes $\{\Delta_t^{(i)},\ t = 1, 2, \ldots\}$ is selected among several ARMA models and different degrees of homogeneity between the genders. In general, the processes for males and females may neither be independent nor be identically distributed. It can even be the case that both processes follow different ARMA dynamics. Empirical data however suggests that there is some similarity between them, as it will be seen in Section 3.

# 3 Application to Belgian mortality data

## 3.1 Presentation of the data

We consider mortality data for Belgian males and females available from Statistics Belgium (`http://statbel.fgov.be/`). The data set considered here consists of ages $x_1 = 65$ to $x_n = 99$ ($n = 35$) observed in the time period from 1970 to 2010. Thus, $t = 1$ corresponds to the mortality improvement from calendar year 1970 to 1971 whereas $T$ corresponds to that from 2009 to 2010. In a later stage of our analysis, we supplement these observations with calendar years 2011 and 2012 to study the stability over successive forecasts.

Figure 3.1 displays the observed age-aggregated mortality improvements $r_{\bullet t}^{(i)}$ for males and females. Both series appear to be strongly correlated. Mortality statistics depicted in Figure 3.1 indicate negative correlation between $r_{\bullet t}^{(i)}$ and $r_{\bullet t+1}^{(i)}$. This property is a consequence of the typical zig-zag pattern, i.e. large improvements in mortality rates are followed by small improvements and vice versa. This apparent behavior also rules out time-invariant random effects in (2.4), i.e. $\Delta_t^{(i)} = \Delta^{(i)}$, as this specification implies $\mathbb{C}\text{ov}[r_{\bullet t}^{(i)}, r_{\bullet s}^{(i)}] = \mathbb{V}\text{ar}[\Delta^{(i)}] > 0$ for all $t \neq s$. Hence, $\Delta_t^{(i)} = \Delta^{(i)}$ constraints $r_{\bullet t}^{(i)}$ and $r_{\bullet s}^{(i)}$ to be positively correlated among all years $t$ and $s$ which contradicts empirical evidence in Figure 3.1

## 3.2 Model selection

As our models are fully specified, with Normally distributed components, the maximum likelihood approach is expected to deliver accurate estimations. However, maximizing the likelihood function associated to the observed age- and gender-specific mortality improvement rates $r_{xt}^{(i)}$, $x \in \{x_1, \ldots, x_n\}$, $t \in \{1, \ldots, T\}$ and $i \in \{m, f\}$ appears to be difficult from a numerical point of view. This observation vector is of length $2nT$ and its covariance matrix of dimension $(2nT) \times (2nT)$ has to be inverted each time the likelihood function is evaluated. Furthermore, optimization involves a large number of parameters, including $2n$ parameters $\beta_x^{(i)}$, the drift parameters $\delta_i$, the time index volatilities $\sigma_{i\Delta}^2$, the correlation parameters $\rho_{ih}$, and the residual variances $\sigma_{i\bullet}^2$.

This motivates to resort first to the aggregate model (2.12) without gender-specific sentivity parameters $\beta_x^{(i)}$. In this model both the dimension of the covariance matrix and the number of parameters are drastically reduced and this greatly facilitates the numerical evaluations.

Let $\boldsymbol{r}$ gather the observed aggregate improvement rates, i.e.

$$\boldsymbol{r} = (r_{\bullet 1}^{(m)}, \ldots, r_{\bullet T}^{(m)}, r_{\bullet 1}^{(f)}, \ldots, r_{\bullet T}^{(f)})'.$$

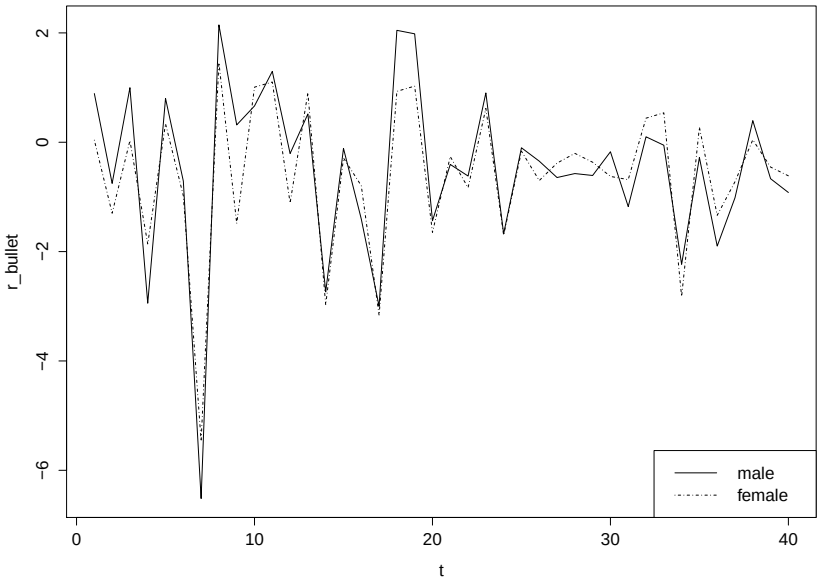


Figure 3.1: Age-aggregated mortality improvements $r_{\bullet t}^{(m)}$ for Belgian males and $r_{\bullet t}^{(f)}$ for females.

In order to obtain the maximum likelihood estimates of the parameters, the log-likelihood function

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

of a $2T$-variate Normal distribution with mean vector $\boldsymbol{\delta}_\bullet$ and variance-covariance matrix $\boldsymbol{\Sigma}_\bullet$ has to be maximized.

The model selection procedure follows a backward approach. We start from a first model allowing for dynamics specific to each gender and we simplify it step by step, as long as the goodness of fit is not significantly worse, as measured by the likelihood ratio test when models are nested.

### 3.2.1 Model 1

We first fit ARMA models for both genders separately, i.e. we model $(r_{\bullet 1}^{(m)}, \ldots, r_{\bullet T}^{(m)})$ and $(r_{\bullet 1}^{(f)}, \ldots, r_{\bullet T}^{(f)})$ in isolation following the Box-Jenkins methodology. Although the marginal models face the identifiability problem described earlier and maximum likelihood estimates are not unique, the AIC values can still be taken for comparison.

We restrict our analysis to ARMA$(p, q)$ models with $0 \leq p + q \leq 5$. This covers the previously mentioned AR(1) structure characterized by the exponentially-decaying correlations (2.8). Models of higher orders are not shown here since additional parameters were not significant and the corresponding AIC values did not show any improvement over those of the lower-order ARMA models. Tables 3.1 and 3.2 display the corresponding AIC values for

| $q\backslash p$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 149.4639 | 148.9936 | 149.3363 | 151.1968 | 150.1263 | 151.8344 |
| 1 | 148.0572 | 146.0322 | 147.7353 | 149.6167 | 147.6900 | |
| 2 | 147.4939 | 147.9160 | 149.7127 | 151.4503 | | |
| 3 | 149.3620 | 149.9059 | 151.6964 | | | |
| 4 | **144.4445** | 146.0678 | | | | |
| 5 | 146.1746 | | | | | |

Table 3.1: AIC values for marginal ARMA modeling of $(r_{\bullet 1}^{(m)}, \ldots, r_{\bullet T}^{(m)})$, males. The optimal value is printed in bold.

males and females. Minimum AIC values are attained at MA(4) for both males and females, i.e. the marginal optimal models are of the form

$$\Delta_t^{(i)} = \delta_i + \theta_{i1}\eta_{t-1}^{(i)} + \theta_{i2}\eta_{t-2}^{(i)} + \theta_{i3}\eta_{t-3}^{(i)} + \theta_{i4}\eta_{t-4}^{(i)} + \eta_t^{(i)}, \;\; t = 1, 2, \ldots, \;\; i \in \{m, f\},$$

with mutually independent errors $\eta_t^{(i)} \sim \mathcal{N}or(0, \sigma_{i\eta}^2)$.

Note that the autoregressive and moving average parameters are optimized instead of the $\rho_h$ entering (2.6)-(2.7). For instance, MA($q$) processes specified by moving average parameters $\theta_0 = 1$ and $\theta_1, \ldots, \theta_q \in \mathbb{R}$ imply correlation parameters

$$\rho_h = \frac{\tilde{\rho}_h}{\tilde{\rho}_0} \text{ with } \tilde{\rho}_h = \begin{cases} \sum_{j=0}^{q-h} \theta_j \theta_{j+h} & h \leq q \\ 0 & h > q. \end{cases}$$

Furthermore, $\sigma_{i\eta}^2$ and $\sigma_{i\Delta}^2$ are equivalent in the sense that both are implicitly given through the relation

$$\sigma_{i\Delta}^2 = \tilde{\rho}_0 \sigma_{i\eta}^2.$$

In addition to these marginal MA(4) fits, we re-estimate all model parameters in a single, gender-combined model ensuring identifiability. In this model, the white noise terms $\eta_t^{(m)}$ and $\eta_t^{(f)}$ are correlated with parameter $\gamma \neq 0$. Precisely, the covariance structure is given by

$$\mathbb{C}\text{ov}[\eta_s^{(m)}, \eta_t^{(f)}] = \begin{cases} 0 & \text{if } s \neq t \\ \gamma \sigma_{m\eta} \sigma_{f\eta} & \text{if } s = t. \end{cases}$$

The joint optimization under the gender-combined structure yields an AIC value of 206.61. The fitted model is described in Table 3.7.

### 3.2.2 Model 2

Let us now assume that $r_{\bullet t}^{(m)}$ and $r_{\bullet t}^{(f)}$ follow the same ARMA(p,q) dynamics with common ARMA parameters. Several ARMA models are tested by maximizing the likelihood of the joint series of gender-specific aggregate improvement factors, correlated through the $\gamma$ parameter. The various models are evaluated by means of their respective AIC. In doing so,

| $q\backslash p$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 136.7372 | 132.7384 | 132.6566 | 131.6732 | 133.5554 | 135.5036 |
| 1 | 131.8829 | 132.3190 | 133.7849 | 133.4047 | 135.3406 | |
| 2 | 133.3640 | 133.2898 | 135.7668 | 134.7513 | | |
| 3 | 133.8895 | 134.0904 | 137.6558 | | | |
| 4 | **131.6661** | 133.3136 | | | | |
| 5 | 133.4416 | | | | | |

Table 3.2: AIC values for marginal ARMA modeling of $(r^{(f)}_{\bullet 1}, \ldots, r^{(f)}_{\bullet T})$, females. The optimal value is printed in bold.

| $q\backslash p$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 213.9461 | 210.3904 | 210.6470 | 211.9065 | 210.8320 | 212.5491 |
| 1 | 208.8653 | 207.9097 | 210.4904 | 212.2603 | 207.6476 | |
| 2 | 208.7503 | 209.8372 | 212.3831 | 214.2449 | | |
| 3 | 210.5695 | 211.3642 | 213.8296 | | | |
| 4 | **206.6793** | 208.7297 | | | | |
| 5 | 208.6672 | | | | | |

Table 3.3: AIC values for joint ARMA modeling, males and females combined. The optimal value is printed in bold.

parameters $\delta_i$, $\sigma^2_{i\Delta}$ and $\sigma^2_{i\bullet}$ are allowed to be gender-specific. The AIC values are displayed in Table 3.3. We can see there that the MA(4) structure is still optimal, with an AIC of 206.68 close to the preceding 206.61. The model fit is described in Table 3.7.

### 3.2.3 Model 3

As the estimated $\gamma$ in Model 2 is close to 1, we now consider a fixed gender-correlation parameter $\gamma = 1$. The AIC values are displayed in Table 3.4. The optimal model is again MA(4) and setting $\gamma = 1$ has little effect on AIC, which changes from 206.68 to 205.85. The estimated model is described in Table 3.7.

### 3.2.4 Model 4

Now, Model 2 is fitted with gender-common parameters $\delta$, $\sigma^2_\Delta$ and $\sigma^2_\bullet$. The results are summarized in Table 3.5 where the optimal AIC value corresponds to the MA(4) model. The estimations of model parameters are displayed in Table 3.7.

### 3.2.5 Model 5

We now set $\gamma$ equal to 1 in Model 4. This is the particular case $\Delta^{(m)}_t = \Delta^{(f)}_t$ in which the mortality improvement factors applying to males and females are both functions of a single

| $q\backslash p$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 211.9461 | 207.9474 | 209.2672 | 210.9794 | 208.6188 | 213.7698 |
| 1 | 207.0918 | 206.3723 | 210.1951 | 210.2307 | 206.2988 | |
| 2 | 207.5804 | 208.2074 | 211.5257 | 212.2713 | | |
| 3 | 208.6789 | 209.2074 | 212.8785 | | | |
| 4 | **205.8454** | 207.5354 | | | | |
| 5 | 207.5295 | | | | | |

Table 3.4: AIC values for joint ARMA modeling, males and females combined, $\gamma = 1$. The optimal value is printed in bold.

| $q\backslash p$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 215.6965 | 213.6938 | 214.0397 | 215.3038 | 215.1425 | 217.3831 |
| 1 | 212.4887 | 210.4227 | 212.9767 | 215.2974 | 209.9891 | |
| 2 | 210.9422 | 212.3710 | 214.9879 | 217.0481 | | |
| 3 | 212.8919 | 214.1754 | 216.3388 | | | |
| 4 | **209.0101** | 211.0094 | | | | |
| 5 | 210.9071 | | | | | |

Table 3.5: AIC values for joint ARMA modeling, males and females combined, with gender-common $\delta$, $\sigma^2_\Delta$ and $\sigma^2_\bullet$. The optimal value is printed in bold.

| $q\backslash p$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 213.6965 | 211.6938 | 212.6375 | 214.3208 | 212.6175 | 214.3594 |
| 1 | 210.4887 | 208.5540 | 211.6871 | 213.3390 | 208.4638 | |
| 2 | 209.6065 | 210.4483 | 213.5879 | 215.1178 | | |
| 3 | 211.0579 | 212.0723 | 215.5870 | | | |
| 4 | **207.7992** | 209.6640 | | | | |
| 5 | 209.6414 | | | | | |

Table 3.6: AIC values for joint ARMA modeling, males and females combined, with $\gamma = 1$ and gender-common $\delta$, $\sigma^2_\Delta$ and $\sigma^2_\bullet$. The optimal value is printed in bold.

| Parameters | Model 1 | | Model 2 | | Model 3 | | Model 4 | | Model 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | m | f | m | f | m | f | m | f | m | f |
| | common | | common | | common | | common | | common | |
| $\theta_1$ | $-0.665$ | $-0.598$ | $-0.516$ | | $-0.496$ | | $-0.479$ | | $-0.487$ | |
| $\theta_2$ | $-0.373$ | $-0.114$ | $-0.254$ | | $-0.232$ | | $-0.279$ | | $-0.251$ | |
| $\theta_3$ | 0.135 | 0.116 | 0.190 | | 0.139 | | 0.183 | | 0.159 | |
| $\theta_4$ | $-0.336$ | $-0.233$ | $-0.417$ | | $-0.408$ | | $-0.396$ | | $-0.420$ | |
| $\delta$ | $-0.503$ | $-0.604$ | $-0.503$ | $-0.604$ | $-0.503$ | $-0.604$ | $-0.554$ | | $-0.554$ | |
| $\sigma^2_\Delta$ | 2.720 | 2.028 | 2.287 | 1.619 | 2.121 | 1.569 | 1.974 | | 1.830 | |
| $\sigma^2_\bullet$ | 0.020 | 0.028 | 0.114 | 0.102 | 0.194 | 0.114 | 0.132 | | 0.180 | |
| $\gamma$ | 0.956 | | 0.978 | | - | | 0.974 | | - | |

Table 3.7: Parameter estimates for Models 1-5.

$\Delta_t$, i.e.

$$r^{(i)}_{xt} = \beta^{(i)}_x \Delta_t + \epsilon^{(i)}_{xt} \quad \text{for } i \in \{m, f\}.$$

The results are summarized in Table 3.6. The fit of the optimal model is described in Table 3.7.

### 3.2.6 Comparison

Let us now compare all nested models by means of likelihood-ratio tests. The $p$-values of the likelihood ratio tests are displayed in Table 3.8. Simple models should be preferred if the difference in log-likelihoods is not significant. As the table shows, the AIC-best Model 3 can not be rejected when compared to Models 1 and 2. Moreover, gender-common Models 4 and 5 are not favored since the corresponding $p$-values are smaller than 5% (except one equal to 5.81%).

| Model | versus Model | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 2 | 0.0892 | - | - | - |
| 3 | 0.1002 | 0.2802 | - | - |
| 4 | 0.0217 | 0.0396 | - | - |
| 5 | 0.0282 | 0.0581 | 0.0469 | 0.3743 |

Table 3.8: Model comparisons: $p$-values of likelihood ratio tests for nested models.

Estimates for the time correlation parameters $\rho_h$ are given by

$$\begin{aligned}\hat{\rho}_1 &= -0.3164\\ \hat{\rho}_2 &= -0.1389\\ \hat{\rho}_3 &= 0.2301\\ \hat{\rho}_4 &= -0.2747\end{aligned}$$

and $\hat{\rho}_h = 0$ for $h \geq 5$. Comparing the autocovariance functions displayed in Figure 3.2, we see that the additional autoregressive parameters do not bring any major change. The optimal gender-common MA(4) specification accurately reproduces the empirical correlations, especially their zig-zag pattern. This legitimates the choice of the AIC-best Model 3 for further analysis of Belgian data. In this case, the state-processes have the structures

$$\begin{aligned}\Delta_t^{(m)} &= \delta_m + \theta_1\eta_{t-1}^{(m)} + \theta_2\eta_{t-2}^{(m)} + \theta_3\eta_{t-3}^{(m)} + \theta_4\eta_{t-4}^{(m)} + \eta_t^{(m)},\\ \Delta_t^{(f)} &= \delta_f + \theta_1\eta_{t-1}^{(f)} + \theta_2\eta_{t-2}^{(f)} + \theta_3\eta_{t-3}^{(f)} + \theta_4\eta_{t-4}^{(f)} + \eta_t^{(f)},\end{aligned}$$

where the error terms $\eta_t^{(i)} \sim \mathcal{N}or(0, \sigma_{i\eta}^2)$ are mutually independent in $t$ for fixed $i \in \{m, f\}$, the pairs $(\eta_t^{(m)}, \eta_t^{(f)})$ being perfectly correlated, i.e.

$$\frac{\eta_t^{(m)}}{\sigma_{m\eta}} = \frac{\eta_t^{(f)}}{\sigma_{f\eta}}$$

holds for all $t$. This ensures that

$$\big(\Delta_t^{(f)} - \delta_f\big) = \frac{\sigma_{f\eta}}{\sigma_{m\eta}}\big(\Delta_t^{(m)} - \delta_m\big),$$

which shows that the gender-specific time indices are perfectly dependent.

## 3.3 Age-specific structure

Given the parameters of the age-aggregate model, we can calibrate the age-specific coefficients $\beta_x^{(i)}$ and the residual variances $\sigma_{i\epsilon}^2$. Notice that for each age $x$, the random vector $(r_{x1}^{(i)}, \ldots, r_{xT}^{(i)})$ is multivariate Normal with mean vector

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

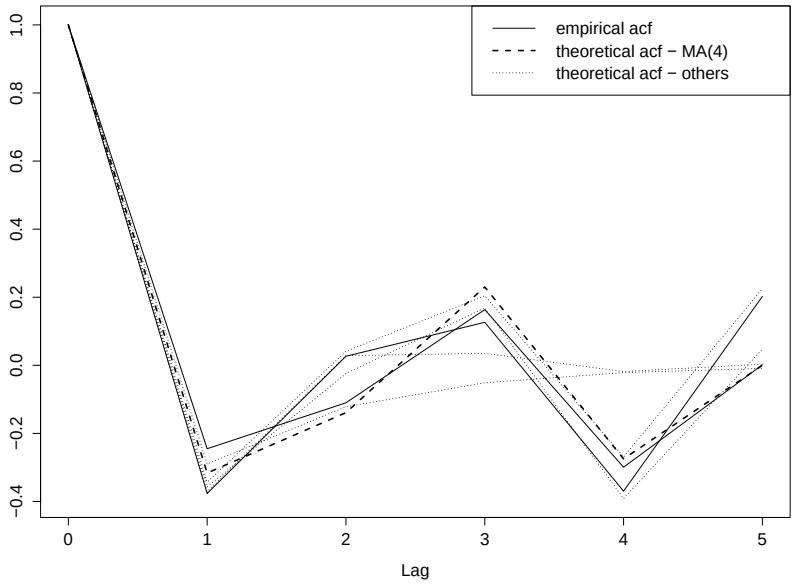


Figure 3.2: The empirical ACF for $r_{\bullet t}^{(m)}$ and $r_{\bullet t}^{(f)}$ and the theoretical ACF of selected ARMA-models.

and variance-covariance matrix

$$\sigma_{i\epsilon}^2 \boldsymbol{I}_T + \beta_x^2 \sigma_{i\Delta}^2 \boldsymbol{C}_T.$$

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

$$\widehat{\delta}_i = \frac{1}{T}\sum_{t=1}^{T} r_{\bullet t}^{(i)}, \tag{3.1}$$

we finally obtain

$$\widehat{\beta}_x^{(i)} = \left(\sum_{t=1}^{T} r_{\bullet t}^{(i)}\right)^{-1} \sum_{t=1}^{T} r_{xt}^{(i)}. \tag{3.2}$$

which sum to 1.

Notice that the estimator (3.2) for $\beta_x^{(i)}$ does not depend on the underlying correlation structure. The estimated parameters are displayed in Figure 3.3 together with their smoothing splines. For $\sigma_{i\epsilon}^2$, we obviously have

$$\widehat{\sigma}_{i\epsilon}^2 = \frac{1}{n}\widehat{\sigma}_{i\bullet}^2. \tag{3.3}$$

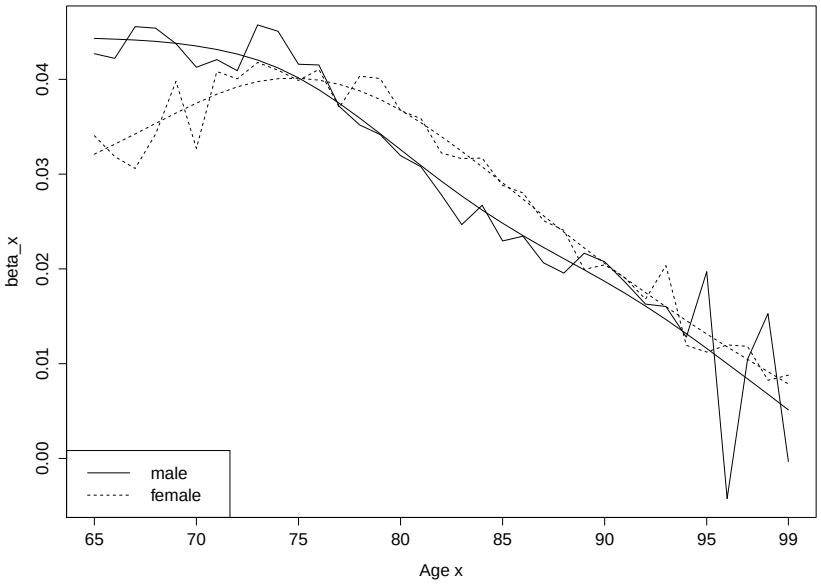


Figure 3.3: Estimated age effects $\beta_x^{(i)}$

# 4 Mortality forecasting

## 4.1 Predictive distribution

In this section, we assume that $\{\Delta_t^{(m)},\ t = 1, 2, \ldots\}$ and $\{\Delta_t^{(f)},\ t = 1, 2, \ldots\}$ are ARMA processes of the same order with gender-specific parameters $\delta_i$, $\sigma^2_{i\Delta}$ and $\sigma^2_{i\bullet}$. The error terms $\epsilon^{(i)}_{\bullet t}$ are independent and identically distributed. This corresponds to Model 2 in Section 3. The predictive distributions for Models 3 to 5 are then easily obtained by making the parameters gender-common or setting $\gamma$ to 1.

Considering (2.6), this formula applies for $\Delta_t^{(m)}$ and $\Delta_t^{(f)}$, replacing $\sigma^2_\Delta$ with $\sigma^2_{m\Delta}$ and $\sigma^2_{f\Delta}$, respectively. Moving to time indices for males and females, we now have covariances

$$\mathbb{C}\mathrm{ov}[\Delta_t^{(m)}, \Delta_s^{(f)}] = \gamma \rho_{|t-s|} \sigma^2_\Delta \tag{4.1}$$

where

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

Rewriting (4.1) in matrix notation, the variance-covariance matrix between the vectors $(\Delta_1^{(m)}, \ldots, \Delta_T^{(m)})'$ and $(\Delta_1^{(f)}, \ldots, \Delta_T^{(f)})'$ is $\gamma\sigma^2_\Delta \boldsymbol{C}_T$, where $\boldsymbol{C}_T$ has been defined in (2.7).

Then, the gender-combined random vector $\boldsymbol{r}$ of past observed aggregate improvement rates is multivariate Normal with mean vector

$$\boldsymbol{\delta}_\bullet = (\delta_m, \ldots, \delta_m, \delta_f, \ldots, \delta_f)' \tag{4.2}$$

and variance-covariance matrix

$$\boldsymbol{\Sigma}_\bullet = \begin{pmatrix} \sigma_{m\bullet}^2 \boldsymbol{I}_T + \sigma_{m\Delta}^2 \boldsymbol{C}_T & \gamma\sigma_\Delta^2 \boldsymbol{C}_T \\ \gamma\sigma_\Delta^2 \boldsymbol{C}_T & \sigma_{f\bullet}^2 \boldsymbol{I}_T + \sigma_{f\Delta}^2 \boldsymbol{C}_T \end{pmatrix}. \tag{4.3}$$

The additional parameter $\gamma$ now solves the identifiability problem of $\boldsymbol{\Sigma}_\bullet$. More precisely, a modification of $\sigma_{i\Delta}^2$ can be compensated by adjusting $\rho_h$ and $\sigma_{i\bullet}^2$ on the diagonal blocks $\sigma_{i\bullet}^2 \boldsymbol{I}_T + \sigma_{i\Delta}^2 \boldsymbol{C}_T$ but not on the off-diagonal blocks $\gamma\sigma_\Delta^2 \boldsymbol{C}_T$. To establish identifiability, we require $\gamma \neq 0$.

The case $\gamma = 1$ is of particular interest as Carter and Lee (1992) suggested using the same time index $\kappa_t$ for both genders. To avoid long-run divergence in gender-specific mortality forecasts, Li and Lee (2005) further proposed to use the same $\beta_x$ and the same $\kappa_t$ for all groups. Here, we nevertheless allow for gender-specific sensitivities $\beta_x^{(i)}$. The situation $\gamma = 1$ prohibits all diversification between mortality improvements for males and females.

In actuarial applications, we are interested in the predictive distribution of the future $\Delta_{T+1}^{(m)}, \ldots, \Delta_{T+k}^{(m)}, \Delta_{T+1}^{(f)}, \ldots, \Delta_{T+k}^{(f)}$ given the past observed aggregate mortality improvement factors $r_{\bullet 1}^{(m)}, \ldots, r_{\bullet T}^{(m)}, r_{\bullet 1}^{(f)}, \ldots, r_{\bullet T}^{(f)}$. The derivation of this predictive distribution requires the distribution of the random vector

$$(r_{\bullet 1}^{(m)}, \ldots, r_{\bullet T}^{(m)}, r_{\bullet 1}^{(f)}, \ldots, r_{\bullet T}^{(f)}, \Delta_{T+1}^{(m)}, \ldots, \Delta_{T+k}^{(m)}, \Delta_{T+1}^{(f)}, \ldots, \Delta_{T+k}^{(f)})' \tag{4.4}$$

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

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

Further, define $\boldsymbol{C}_{k,T} = \boldsymbol{C}_{T,k}'$. Therefore, the random vector (4.4) is multivariate Normal with mean vector $\delta \mathbf{1}_{2T+2k}$ and variance-covariance matrix

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

Therefore, given $(r_{\bullet 1}^{(m)}, \ldots, r_{\bullet T}^{(m)}, r_{\bullet 1}^{(f)}, \ldots, r_{\bullet T}^{(f)})'$, the predictive distribution for the future $(\Delta_{T+1}^{(m)}, \ldots, \Delta_{T+k}^{(m)}, \Delta_{T+1}^{(f)}, \ldots, \Delta_{T+k}^{(f)})'$ is multivariate Normal with mean vector

$$\begin{pmatrix} \delta_m \mathbf{1}_k \\ \delta_f \mathbf{1}_k \end{pmatrix} + \begin{pmatrix} \sigma_{m\Delta}^2 \boldsymbol{C}_{k,T} & \gamma\sigma_\Delta^2 \boldsymbol{C}_{k,T} \\ \gamma\sigma_\Delta^2 \boldsymbol{C}_{k,T} & \sigma_{f\Delta}^2 \boldsymbol{C}_{k,T} \end{pmatrix} \begin{pmatrix} \sigma_{m\bullet}^2 \boldsymbol{I}_T + \sigma_{m\Delta}^2 \boldsymbol{C}_T & \gamma\sigma_\Delta^2 \boldsymbol{C}_T \\ \gamma\sigma_\Delta^2 \boldsymbol{C}_T & \sigma_{f\bullet}^2 \boldsymbol{I}_T + \sigma_{f\Delta}^2 \boldsymbol{C}_T \end{pmatrix}^{-1} \begin{pmatrix} r_\bullet^{(m)} - \delta_m \mathbf{1}_T \\ r_\bullet^{(f)} - \delta_f \mathbf{1}_T \end{pmatrix} \tag{4.5}$$

and variance-covariance matrix

$$\begin{pmatrix} \sigma^2_{m\Delta}\boldsymbol{C}_k & \gamma\sigma^2_{\Delta}\boldsymbol{C}_k \\ \gamma\sigma^2_{\Delta}\boldsymbol{C}_k & \sigma^2_{f\Delta}\boldsymbol{C}_k \end{pmatrix} - \begin{pmatrix} \sigma^2_{m\Delta}\boldsymbol{C}_{k,T} & \gamma\sigma^2_{\Delta}\boldsymbol{C}_{k,T} \\ \gamma\sigma^2_{\Delta}\boldsymbol{C}_{k,T} & \sigma^2_{f\Delta}\boldsymbol{C}_{k,T} \end{pmatrix}$$

$$\begin{pmatrix} \sigma^2_{m\bullet}\boldsymbol{I}_T + \sigma^2_{m\Delta}\boldsymbol{C}_T & \gamma\sigma^2_{\Delta}\boldsymbol{C}_T \\ \gamma\sigma^2_{\Delta}\boldsymbol{C}_T & \sigma^2_{f\bullet}\boldsymbol{I}_T + \sigma^2_{f\Delta}\boldsymbol{C}_T \end{pmatrix}^{-1} \begin{pmatrix} \sigma^2_{m\Delta}\boldsymbol{C}_{T,k} & \gamma\sigma^2_{\Delta}\boldsymbol{C}_{T,k} \\ \gamma\sigma^2_{\Delta}\boldsymbol{C}_{T,k} & \sigma^2_{f\Delta}\boldsymbol{C}_{T,k} \end{pmatrix}. \tag{4.6}$$

The predictive distributions of $(r^{(i)}_{\bullet T+1}, \ldots, r^{(i)}_{\bullet T+k})$ and $(r^{(i)}_{x,T+1}, \ldots, r^{(i)}_{x,T+k})$ then easily follow.

## 4.2 Mortality forecast

Given the predictive distribution for the future $\Delta^{(i)}_{T+k}$, one can obtain forecasts for the future death rates $m^{(i)}_{x,T+k}$ and the corresponding one-year death probabilities $q^{(i)}_{x,T+k}$. In detail, by iterating the relationship

$$m^{(i)}_x(t) = m^{(i)}_x(t-1)\exp\left(\beta^{(i)}_x\Delta^{(i)}_t + \epsilon^{(i)}_{xt}\right),$$

we get

$$m^{(i)}_x(T+k) = m^{(i)}_x(T)\exp\left(\sum_{j=1}^{k}\left(\beta^{(i)}_x\Delta^{(i)}_{T+j} + \epsilon^{(i)}_{x,T+j}\right)\right). \tag{4.7}$$

To get point predictions of $m^{(i)}_x(T+k)$, we can use median death rates

$$\widehat{m}^{(i)}_x(T+k) = m^{(i)}_x(T)\exp\left(\sum_{j=1}^{k}\widehat{\beta}^{(i)}_x\mathbb{E}\left[\Delta^{(i)}_{T+j} \mid r^{(m)}_{\bullet 1}, \ldots, r^{(m)}_{\bullet T}, r^{(f)}_{\bullet 1}, \ldots, r^{(f)}_{\bullet T}\right]\right) \tag{4.8}$$

where the conditional expectations follow from (4.5).

Paths of future $m^{(i)}_x(T+k)$ can be simulated by (4.7). The corresponding one-year death probabilities $q^{(i)}_{x,T+k}$ and one-year survival probabilities $p^{(i)}_{x,T+k}$ are easily obtained from

$$\widehat{q}^{(i)}_{x,T+k} = 1 - \widehat{p}^{(i)}_{x,T+k} = 1 - \exp\left(-\widehat{m}^{(i)}_x(T+k)\right).$$

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

## 4.3 Period life expectancies

We illustrate the forecasts of our mortality model on the basis of period life expectancies. Using the predictive distribution (4.5)-(4.6) of the future $\Delta^{(i)}_{T+k}$ and the procedure described in Subsection 4.2, predictions for future mortality rates are derived and the predicted period life expectancy $\hat{e}_{65}(T+k)$ at age 65 in calendar year $T+k$ can then be calculated using the formula

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

The predicted mortality improvements are applied on the last observation $m_x(2010)$ through (4.8). Table 4.2 shows point predictions $\hat{e}_{65}(2050)$ using formula (4.8). Point forecasts

| | male | male (LC) | female | female (LC) |
|---|---|---|---|---|
| $\delta_i$ | $-0.5037$ | $-0.4345$ | $-0.6045$ | $-0.5221$ |
| $\sigma^2_{i\Delta}$ | 2.1213 | 0.3299 | 1.5692 | 0.6637 |
| $\sigma^2_{i\bullet}$ | 0.1945 | 1.7904 | 0.1149 | 0.6994 |

Table 4.1: Estimated mean and variance parameters. For the Lee-Carter models, the rows show estimates for $\delta_{i\kappa}$, $\sigma^2_{i\kappa}$ and $\sigma^2_{i\circ}$ respectively.

obtained by the Lee-Carter model are shown for comparison. We see that smoothing the estimated age effects $\beta_x^{(i)}$ has little impact on the life expectancy. 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, all calculations are done with smoothed $\beta_x^{(i)}$ and Lee-Carter forecasts use $m_x(T)$ as an initial value instead of the offset $\alpha_x$. In this case, the forecasts roughly agree.

Table 4.1 displays estimates for the age-aggregate mortality improvement model implicitly given by the Lee-Carter model. More precisely, (2.3) gives

$$\sum_{x=x_1}^{x_n} \big( \ln m_x^{(i)}(t) - \ln m_x^{(i)}(t-1) \big) = \sum_{x=x_1}^{x_n} \beta_x^{(i)} (\kappa_t^{(i)} - \kappa_{t-1}^{(i)}) = \kappa_t^{(i)} - \kappa_{t-1}^{(i)} \text{ for } i \in \{m, f\}.$$

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

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

with independent error components $\xi_t^{(i)} \sim \mathcal{N}or(0, \sigma^2_{i\kappa})$. Furthermore, the residual variance between the observed and fitted model is denoted by $\sigma^2_{i\circ}$ which is the analog term to $\sigma^2_{i\bullet}$ in our model. Even though both models are based on the same 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 variance is allocated by the two models. While the Lee-Carter mortality improvement model gives more weight to the measurement variance $\sigma^2_{i\circ}$, the innovation variance $\sigma^2_{i\Delta}$ is dominating in our model. Notice that the innovation error affects all ages through the sensitivity factor $\beta_x^{(i)}$.

Next, $e_{65}(2050)$ has been calculated for 1000 scenarios of simulated life tables for year 2050. To stress the role of the age-common processes $\Delta_t^{(i)}$ and their Lee-Carter counterparts $(\kappa_t^{(i)} - \kappa_{t-1}^{(i)})$, the noise terms $\epsilon_{xt}$ have been set to zero. Table 4.3 lists the empirical means and standard deviations. Although $\Delta_t^{(i)}$ have larger standard deviations than their Lee-Carter counterparts, the opposite is the case for the life expectancies. This might be counter-intuitive at first sight but it is a consequence of the underlying moving-average structure. As the estimated autocorrelation function of the time index is negative for lags of size one, large deviations of $\Delta_t^{(i)}$ are likely to followed by $\Delta_{t+1}^{(i)}$ going into the opposite direction. By (4.7), the deviations cancel out. On the other hand, mortality improvements are independent under the Lee-Carter model. Thus, outliers remain and strongly impact the future life expectancy.

| | MA(4) | MA(4) with smoothed $\beta_x$ | LC with $m_x(2010)$ | LC with $\alpha_x$ |
|---|---|---|---|---|
| Males | 22.1922 | 22.2143 | 21.9329 | 19.5003 |
| Females | 25.7450 | 25.7639 | 25.3079 | 23.2666 |

Table 4.2: Point forecasts of period life expectancy in 2050

| | Males | Males (LC) | Females | Females (LC) |
|---|---|---|---|---|
| Mean | 22.2165 | 21.9196 | 25.7656 | 25.2412 |
| Std | 0.3367 | 0.7140 | 0.2322 | 0.8117 |

Table 4.3: mean and standard deviation obtained from simulated life tables

## 4.4 Stability over successive forecasts

Let us now show that the model proposed in the present paper solves the stability 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. As Table 4.4 shows, our estimates are more stable than the other two. This is again a consequence of the moving average structure, i.e. mortality improvements not being independent in time.

This effect is further illustrated in Figure 4.1. We have displayed there the forecasts for $e_{65}(T+1), \ldots, e_{65}(2017)$ 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.

| | male | male (LC) | male (official) | female | female (LC) | female (official) |
|---|---|---|---|---|---|---|
| up to 2012 | 22.31 | 21.86 | 23.04 | 25.72 | 25.16 | 25.12 |
| up to 2011 | 22.28 | 22.12 | 23.21 | 25.78 | 25.41 | 25.29 |
| up to 2010 | 22.21 | 21.93 | 22.91 | 25.76 | 25.30 | 25.00 |

Table 4.4: Predicted life expectancies for different data sets

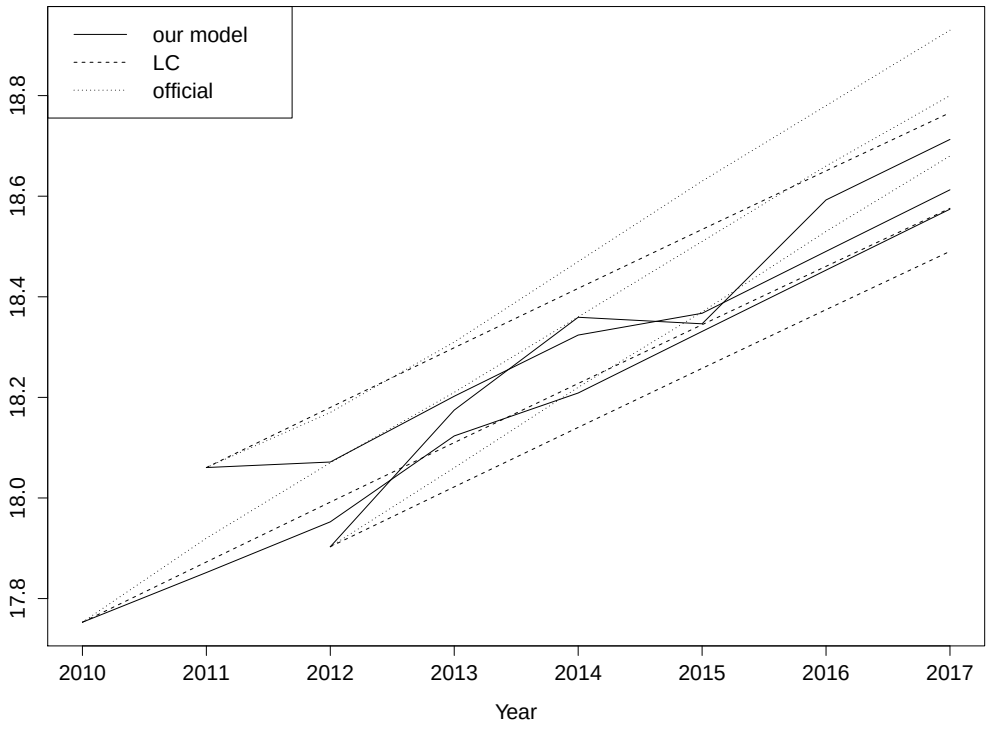


Figure 4.1: Predicted life expectancies for different data sets. For each of the three models, the three consecutive lines are based on data up to 2010, 2011 and 2012 respectively.

# Acknowledgements

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


# References


- Borger, M., Aleksic, M.C., 2013. Coherent projections of age, period, and cohort dependent mortality improvements. Technical report.

- Carter, L., Lee, R.D., 1992. Modeling and forecasting US sex differentials in mortality. International Journal of Forecasting 8, 393-411.

- Czado, C., Delwarde, A., Denuit, M., 2005. Bayesian Poisson log-bilinear mortality projections. Insurance: Mathematics and Economics 36, 260-284.

- Girosi, F., King, G., 2008. Demographic Forecasting. Princeton University Press.

- Kogure, A., Kitsukawa, K., Kurachi, Y., 2009. A Bayesian comparison of models for changing mortalities toward evaluating the longevity risk in Japan. Asia-Pacific Journal of Risk and Insurance 3, 1-22.

- Kogure, A., Kurachi, Y., 2010. A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. Insurance: Mathematics and Economics 46, 162-172.

- Li, N., Lee, R.D., 2005. Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. Demography 42, 575-594.

- Lee, R.D., Carter, L., 1992. Modelling and forecasting the time series of US mortality. Journal of the American Statistical Association 87, 659-671.

- Li, J., 2014. An application of MCMC simulation in mortality projection for populations with limited data. Demographic Research 30, 1-48.

- Mitchell, D., Brockett, P., Mendoza-Arriaga, R., Muthuraman, K., 2013. Modeling and forecasting mortality rates. Insurance: Mathematics and Economics 52, 275-285.

- Pedroza, C., 2006. A Bayesian forecasting model: Predicting U.S. male mortality. Biostatistics 7, 530-550.

- Sundt, B., 1981. Recursive credibility estimation. Scandinavian Actuarial Journal, 3-21.