StMoMo: An R Package for Stochastic Mortality Modeling

Andrés M. Villegas
UNSW Business School
UNSW Sydney

Pietro Millossovich
Cass Business School
City, University of London
DEAM, University of Trieste

Vladimir K. Kaishev
Cass Business School
City, University of London

Abstract

In this paper we mirror the framework of generalized (non-)linear models to define the family of generalized age-period-cohort stochastic mortality models which encompasses the vast majority of stochastic mortality projection models proposed to date, including the well-known Lee-Carter and Cairns-Blake-Dowd models. We also introduce the R package **StMoMo** which exploits the unifying framework of the generalized age-period-cohort family to provide tools for fitting stochastic mortality models, assessing their goodness of fit and performing mortality projections. We illustrate some of the capabilities of the package by performing a comparison of several stochastic mortality models applied to the England and Wales population.

Keywords: mortality modeling; mortality forecasting; age-period-cohort; generalized non-linear models.

1. Introduction

During the last two centuries developed countries experienced a persistent increase in life expectancy. For instance, Oeppen and Vaupel (2002) estimate that during the last 160 years the world record in female life expectancy at birth has increased at an approximate steady pace of 3 months per year. This increase in life expectancy, though a sign of social progress, poses a challenge to governments, private pension plans and life insurers because of its impact on pension and health costs. Actuaries and demographers have recognized the problems caused by an aging population and rising longevity and have thus devoted significant attention to the development of statistical techniques for the modeling and projection of mortality rates.

One of the most influential approaches to the stochastic modeling of mortality rates is the parsimonious mortality model proposed by Lee and Carter (1992). This model uses principal component analysis to decompose the age-time matrix of mortality rates into a bilinear combination of age and period parameters, with the latter being treated as time series to produce mortality projections. The Lee-Carter model has inspired numerous variants and extensions. For instance, Lee and Miller (2001), Booth *et al.* (2002), and Brouhns *et al.* (2002) have proposed alternative estimation approaches in order to improve the goodness-of-fit and the forecasting properties of the model. In particular, Brouhns *et al.* (2002) propose a more formal

statistical approach to estimating the parameters by embedding the Lee-Carter model into a Poisson regression setting. Other authors have extended the Lee-Carter model by including additional terms, such as multiple bilinear age-period components (Renshaw and Haberman 2003; Hyndman and Ullah 2007), or a cohort effect term (Renshaw and Haberman 2006).

The two factor Cairns-Blake-Dowd (CBD) model introduced by Cairns *et al.* (2006) is one of the most prominent variants of the Lee-Carter model. The CBD model relies on the linearity of the logit of one-year death probabilities at older ages. Specifically, it assumes that, for a given, year the logit of the one-year death probability is a linear function of age, and treats the intercept and slope parameters across years as stochastic processes. Cairns *et al.* (2009) consider three extensions to the original CBD model by incorporating combinations of a quadratic age term and a cohort effect term. Plat (2009) has combined features of the CBD and the Lee-Carter models to produce a model that is suitable for full age ranges and captures the cohort effect.

Given the abundance and rapid increase in the number of stochastic mortality models proposed in the literature, there have been some recent attempts to find the commonalities among these models. Hunt and Blake (2015) review the structure of mortality models and describe an age-period-cohort model structure which encompasses the vast majority of stochastic mortality models. Currie (2016) shows that many mortality models can be expressed in terms of generalized linear models or generalized non-linear models.

In this paper, we build upon the works of Hunt and Blake (2015) and Currie (2016) to define the family of generalized age-period-cohort stochastic mortality models by mirroring the terminology of generalized linear models. We also introduce the R package **StMoMo** (Villegas *et al.* 2017) which exploits the unifying framework of the generalized age-period-cohort family combined with the powerful fitting function of the **gnm** package (Turner and Firth 2015) to provide computational tools for implementing many of the stochastic mortality models proposed to date.[1] The **StMoMo** package is available at `http://cran.r-project.org/package=StMoMo`. Version 0.4.0 has been used for this paper.

Several packages for mortality modeling are available in the R environment (R Core Team 2016). The package **demography** (Hyndman *et al.* 2014), whose usage is explained in detail in Booth *et al.* (2014), implements, among other things, the original Lee-Carter model along with the Lee and Miller (2001), Booth *et al.* (2002), and Hyndman and Ullah (2007) variants. The **ilc** package (Butt *et al.* 2014) implements the Renshaw and Haberman (2006) cohort extension of the Lee-Carter model together with the Lee-Carter model under a Poisson regression framework. The **LifeMetrics** R functions implement the original CBD model and the three extended CBD models considered in Cairns *et al.* (2009), along with the Lee-Carter model (using Poisson maximum likelihood), the traditional age-period-cohort model (see, Osmond (1985)) and the Renshaw and Haberman (2006) model. This package, which is not on CRAN, is available at `http://www.macs.hw.ac.uk/~andrewc/lifemetrics/`.

---

[1]The acronym StMoMo, pronounced Saint Momo, stands for Stochastic Mortality Modeling. Momo is the king of Carnivals in numerous Latin American festivities (Wikipedia 2014).

There are however several drawbacks of the existing packages which our package **StMoMo** seeks to overcome. First, the existing packages are based on model-specific fitting algorithms limiting the models available to those already predefined in the packages. By contrast, **StMoMo** allows users to easily expand the number of models available. In addition, whilst **StMoMo** provides forecasting and simulation functions for any model within the generalized age-period-cohort family, the existing packages only provide such functions for a limited number of models. For instance, the package **ilc** only includes forecasting functions for the Lee-Carter model. Similarly, simulation with the package **LifeMetrics** is limited to the Lee-Carter and the standard CBD models. Finally, **StMoMo** provides functions which are not available in existing packages, such as tools for analyzing the goodness-of-fit[2] and evaluating the impact of parameter uncertainty using bootstrapping techniques.

StMoMo comes with a set of functions for defining an abstract model — specifying for instance the number of period terms, whether coefficients are parametric or not — and for fitting a given model. This is particularly useful when estimating several models on a given dataset or a given model to different datasets. The package also provides preset functions for defining the most common models available in the mortality forecasting literature. In addition, other models preferred by the user can be created in a very simple fashion, see Section 4 where several examples are given. Therefore, the flexibility of the package allows a user to quickly build up a battery of different models, and this is particularly useful when seeking the most appropriate mortality model, comparing different models or assessing model risk. **StMoMo** is particularly appealing for actuaries managing life and pensions portfolios exposed to longevity risk. The code backing most functions implemented in the package has been extensively used and tested for the development of multi-population mortality models for assessing basis risk in longevity risk transactions, see Haberman *et al.* (2014).

In this paper we describe the statistical framework underlying **StMoMo** and illustrate its usage. For this purpose, we use as a running example a comparison of several stochastic mortality models fitted to the England and Wales population. This example is in the spirit of the comparison exercises of Cairns *et al.* (2009, 2011), Haberman and Renshaw (2011) and Lovász (2011), allowing us to show how several of the analysis performed in these papers can easily be replicated using **StMoMo**. The structure of the paper is as follows. In Section 2 we introduce our notation. In Section 3, we mirror the framework of generalized linear models to define the family of generalized age-period-cohort (GAPC) stochastic mortality models and demonstrate that many of the mortality models discussed in the literature can be framed within this family. In Section 4 we explain how the GAPC family of models is implemented in **StMoMo**. In Section 5, we describe the fitting of GAPC mortality models and illustrate how this can be accomplished using **StMoMo**. In Section 6 we consider the evaluation of the goodness-of-fit of GAPC models. In Section 7 we discuss the forecasting and simulation of GAPC models using time series techniques. Section 8 describes the use of bootstrapping techniques to incorporate parameter uncertainty in the estimation and forecasting of GAPC mortality models. Finally, in Section 9 we provide some conclusions and discuss possible extensions of the **StMoMo** package.

---

[2]We note that **ilc** also provides some graphical tools for assessing the goodness of fit of the models implemented in that package.

2. Notation and data

Let the random variable D_{xt} denote the number of deaths in a population at age x last birthday during calendar year t. Also let d_{xt} denote the observed number of deaths, E_{xt}^c the central exposed to risk at age x in year t, and E_{xt}^0 the corresponding initial exposed to risk. The one-year death probability for an individual aged x last birthday and in calendar year t, denoted q_{xt}, can be estimated as $\hat{q}_{xt} = d_{xt}/E_{xt}^0$. The force of mortality and central death rates are denoted by μ_{xt} and m_{xt}, respectively, with the empirical estimate of the latter being $\hat{m}_{xt} = d_{xt}/E_{xt}^c$. Under the assumption that the force of mortality is constant over each year of age and calendar year, i.e., from age x to age $x+1$ and year t to $t+1$, then the force of mortality μ_{xt} and the death rate m_{xt} coincide. We assume that this is the case throughout.

In **StMoMo** and throughout this paper we assume that deaths, d_{xt}, and either central exposures, E_{xt}^c, or initial exposures, E_{xt}^0, are available in a rectangular array format comprising ages (on the rows) $x = x_1, x_2, \ldots, x_k$, and calendar years (on the columns) $t = t_1, t_2, \ldots, t_n$, When only central exposures are available and initial exposures are required (or vice-versa), one can approximate the initial exposures by adding half the matching reported numbers of deaths to the central exposures, i.e, $E_{xt}^0 \approx E_{xt}^c + \frac{1}{2}d_{xt}$. When the context is clear, we may write E_{xt} to refer to E_{xt}^0 or E_{xt}^c.

3. Generalized APC stochastic mortality models

Some authors have recently sought to identify the similarities among stochastic mortality models. For instance, Hunt and Blake (2015) describe an age-period-cohort (APC) model structure which encompasses the vast majority of stochastic mortality models. In another interesting contribution, Currie (2016) shows that many common mortality models can be expressed in the standard terminology of generalized linear or non-linear models. In this section, we build upon the aforementioned papers to define the family of generalized age-period-cohort (GAPC) stochastic mortality models.

Akin to generalized linear models (see, e.g., McCullagh and Nelder (1989)), a GAPC stochastic mortality model is comprised of four components:

1. The *random component*: the numbers of deaths D_{xt} follow a Poisson distribution or a Binomial distribution, so that

$$D_{xt} \sim \text{Poisson}(E_{xt}^c \mu_{xt})$$

   or

$$D_{xt} \sim \text{Binomial}(E_{xt}^0, q_{xt}),$$

   with $\mathbb{E}\left(D_{xt}/E_{xt}^c\right) = \mu_{xt}$ and $\mathbb{E}\left(D_{xt}/E_{xt}^0\right) = q_{xt}$, respectively.

2. The *systematic component*: following Hunt and Blake (2015) the effects of age x, calendar year t and year-of-birth (cohort) $c = t - x$ are captured through a *predictor* η_{xt} given by:

$$\eta_{xt} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)}\kappa_t^{(i)} + \beta_x^{(0)}\gamma_{t-x}.$$

Here:

- The term α_x is a static age function capturing the general shape of mortality by age.
- $N \geq 0$ is an integer indicating the number of age-period terms describing the mortality trends, with each time index $\kappa_t^{(i)}$, $i = 1, \ldots, N$, contributing in specifying the mortality trend and $\beta_x^{(i)}$ modulating its effect across ages.
- The term γ_{t-x} accounts for the cohort effect with $\beta_x^{(0)}$ modulating its effect across ages.

The age modulating terms $\beta_x^{(i)}$, $i = 0, 1, \ldots, N$, can be either pre-specified functions of age, i.e., $\beta_x^{(i)} \equiv f^i(x)$, as in CBD type models, or non-parametric terms without any prior structure which need to be estimated as in the Lee-Carter model. In the GAPC family we assume that the period indexes $\kappa_t^{(i)}$, $i = 1, \ldots, N$, and the cohort index γ_{t-x} are stochastic processes. This is the key feature that allows the stochastic projection of GAPC models and thus the generation of probabilistic forecasts of future mortality rates.

3. The *link function g* associating the random component and the systematic component so that

$$g\left(\mathbb{E}\left(\frac{D_{xt}}{E_{xt}} \right) \right) = \eta_{xt}.$$

Although a number of link functions would be possible, it is convenient to use the so-called canonical link and pair the Poisson distribution with the log link function and the Binomial distribution with the logit link function (see, e.g., Currie (2016) for a discussion of this in the context of mortality models and McCullagh and Nelder (1989) in the wider context of GLMs).

4. The *set of parameter constraints*: most stochastic mortality models are only identifiable up to a transformation and thus require parameter constraints to ensure unique parameter estimates. These parameter constraints are applied through a *constraint function v* which maps an arbitrary vector of parameters

$$\theta := \left(\alpha_x, \beta_x^{(1)}, ..., \beta_x^{(N)}, \kappa_t^{(1)}, ..., \kappa_t^{(N)}, \beta_x^{(0)}, \gamma_{t-x} \right)$$

into a vector of transformed parameters

$$v(\theta) = \tilde{\theta} = \left(\tilde{\alpha}_x, \tilde{\beta}_x^{(1)}, ..., \tilde{\beta}_x^{(N)}, \tilde{\kappa}_t^{(1)}, ..., \tilde{\kappa}_t^{(N)}, \tilde{\beta}_x^{(0)}, \tilde{\gamma}_{t-x} \right)$$

satisfying the model constraints with no effect on the predictor η_{xt} (i.e., θ and $\tilde{\theta}$ result in the same η_{xt}).

Most stochastic mortality models proposed in the literature belong to the GAPC family. This includes the original Lee-Carter model, the extensions of the Lee-Carter proposed in

Renshaw and Haberman (2003, 2006), the original CBD model, and the extended CBD models of Cairns *et al.* (2009). In addition, all the model structures considered in Haberman and Renshaw (2011), Lovász (2011) and van Berkum *et al.* (2014), as well as the models of Plat (2009), Aro and Pennanen (2011), O'Hare and Li (2012), Börger *et al.* (2013) and Alai and Sherris (2014), are part of the GAPC class of models.[3]

Next, we describe in detail some of these models highlighting how they can be framed within the GAPC family.

3.1. Lee-Carter model under a Poisson setting

Brouhns *et al.* (2002) have implemented the Lee-Carter model assuming a Poisson distribution of the number of deaths and using the log *link function* with respect to the force of mortality μ_{xt}. The predictor structure proposed by Lee and Carter (1992) assumes that there is a static age function, α_x, a unique non-parametric age-period term ($N = 1$), and no cohort effect. Thus, the predictor is given by:

$$\eta_{xt} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} \tag{1}$$

In order to project mortality, the time index $\kappa_t^{(1)}$ is modeled and forecasted using ARIMA processes. Typically, a random walk with drift has been shown to provide a reasonable fit, that is,

$$\kappa_t^{(1)} = \delta + \kappa_{t-1}^{(1)} + \xi_t, \qquad \xi_t \sim N(0, \sigma_\kappa^2) \qquad \text{i.i.d.,}$$

where δ is the drift parameter and ξ_t is a Gaussian white noise process with variance σ_κ^2.

The Lee-Carter model is only identifiable up to a transformation, as for arbitrary real constants c_1 and $c_2 \neq 0$ the parameters in Equation 1 can be transformed in the following way

$$\left(\alpha_x, \beta_x^{(1)}, \kappa_t^{(1)}\right) \rightarrow \left(\alpha_x + c_1\beta_x^{(1)}, \frac{1}{c_2}\beta_x^{(1)}, c_2(\kappa_t^{(1)} - c_1)\right), \tag{2}$$

leaving η_{xt} unchanged. To ensure identifiability of the model, Lee and Carter (1992) suggest the following *set of parameter constraints*

$$\sum_x \beta_x^{(1)} = 1, \qquad \sum_t \kappa_t^{(1)} = 0, \tag{3}$$

which can be imposed by choosing

$$c_1 = \frac{1}{n}\sum_t \kappa_t^{(1)}, \qquad c_2 = \sum_x \beta_x^{(1)} \tag{4}$$

in transformation (2).

3.2. Renshaw and Haberman model: Lee-Carter with cohort effects

---

[3]We note however that models which rely on the smoothness of mortality over both age and time, such as the graduation approach of Renshaw *et al.* (1996) and the P-Spline model of Currie *et al.* (2006), do not belong to the GAPC family. The P-Spline approach is implemented in the **MortalitySmooth** R package (Camarda 2012).

Renshaw and Haberman (2006) generalize the Lee-Carter model by incorporating a cohort effect to obtain the predictor:

$$\eta_{xt} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(0)}\gamma_{t-x} \tag{5}$$

Mortality projections for this model are derived using time series forecast of the estimated $\kappa_t^{(1)}$ and γ_{t-x}, generated using univariate ARIMA processes under the assumption of independence between the period and the cohort effects.

In order to estimate the model, Renshaw and Haberman (2006) assume a Poisson distribution of deaths (*random component*) and use a log *link* function targeting the force of mortality μ_{xt}. As with the Lee-Carter model, the predictor η_{xt} is invariant to the transformation:

$$\left(\alpha_x, \beta_x^{(1)}, \kappa_t^{(1)}, \beta_x^{(0)}, \gamma_{t-x}\right) \to \left(\alpha_x + c_1\beta_x^{(1)} + c_2\beta_x^{(0)}, \frac{1}{c_3}\beta_x^{(1)},\right.$$
$$\left. c_3(\kappa_t^{(1)} - c_1), \frac{1}{c_4}\beta_x^{(0)}, c_4(\gamma_{t-x} - c_2)\right), \tag{6}$$

where c_1, c_2, $c_3 \neq 0$ and $c_4 \neq 0$ are real constants. Identifiability of the model can be ensured using the following *set of parameter constraints*:

$$\sum_x \beta_x^{(1)} = 1, \qquad \sum_t \kappa_t^{(1)} = 0, \qquad \sum_x \beta_x^{(0)} = 1, \qquad \sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c = 0,$$

which can be imposed by setting

$$c_1 = \frac{1}{n}\sum_t \kappa_t^{(1)}, \quad c_2 = \frac{1}{n+k-1}\sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c, \quad c_3 = \sum_x \beta_x^{(1)}, \quad c_4 = \sum_x \beta_x^{(0)},$$

in transformation (6).

Renshaw and Haberman (2006) also consider several substructures of the predictor (5) obtained by setting to a constant one or both of the age modulating terms. Of particular interest is the substructure obtained by setting $\beta_x^{(0)} = 1$,

$$\eta_{xt} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \gamma_{t-x}, \tag{7}$$

which has been suggested by Haberman and Renshaw (2011) as a simpler structure that resolves some stability issues of the original model.

3.3. APC model

Another commonly used substructure of the Renshaw and Haberman model is the so-called age-period-cohort (APC) model, corresponding to $\beta_x^{(1)} = 1$, $\beta_x^{(0)} = 1$,

$$\eta_{xt} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x},$$

which has a long-standing tradition in the fields of medicine and demography (see, e.g., Clayton and Schifflers (1987), Hobcraft *et al.* (1982) and Osmond (1985)), but has not been

widely used in the actuarial literature until it was considered by Currie (2006). The APC model is known to be invariant with respect to the following two transformations:

$$\left(\alpha_x, \kappa_t^{(1)}, \gamma_{t-x}\right) \rightarrow \left(\alpha_x + \phi_1 - \phi_2 x, \kappa_t^{(1)} + \phi_2 t, \gamma_{t-x} - \phi_1 - \phi_2(t-x)\right) \tag{8}$$

$$\left(\alpha_x, \kappa_t^{(1)}, \gamma_{t-x}\right) \rightarrow \left(\alpha_x + c_1, \kappa_t^{(1)} - c_1, \gamma_{t-x}\right), \tag{9}$$

where c_1, ϕ_1, and ϕ_2 are real constants. However, we can ensure identifiability of the model by imposing the *set of constraints*:

$$\sum_t \kappa_t^{(1)} = 0, \qquad \sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c = 0, \qquad \sum_{c=t_1-x_k}^{t_n-x_1} c\gamma_c = 0,$$

where the last two constraints imply that the cohort effect fluctuates around zero with no discernible linear trend. Following Haberman and Renshaw (2011, Appendix A), the constraints on the cohort effect can be imposed by applying transformation (8) with constants ϕ_1 and ϕ_2 obtained by regressing γ_{t-x} on $t - x$, so that

$$\gamma_{t-x} = \phi_1 + \phi_2(t-x) + \epsilon_{t-x}, \quad \epsilon_{t-x} \sim N(0, \sigma^2) \quad \text{i.i.d.}.$$

The constraint on the period index can then be imposed by applying transformation (9) with

$$c_1 = \frac{1}{n} \sum_t \kappa_t^{(1)}.$$

3.4. CBD model

Cairns *et al.* (2006) propose a predictor structure with two age-period terms ($N = 2$) with pre-specified age-modulating parameters $\beta_x^{(1)} = 1$ and $\beta_x^{(2)} = x - \bar{x}$, no static age function and no cohort effect. Thus, the predictor of the CBD model is given by:

$$\eta_{xt} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)},$$

where $\bar{x}$ is the average age in the data. Cairns *et al.* (2006) obtain mortality forecasts by projecting the period effects $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ using a bivariate random walk with drift.

The CBD model does not have identifiability issues and hence the *set of parameter constraints* is empty. In order to estimate the parameter of the CBD model we can follow Haberman and Renshaw (2011) and assume a Binomial distribution of deaths using a logit *link function* targeting the one-year death probabilities q_{xt}.

3.5. M7: Quadratic CBD model with cohort effects

Cairns *et al.* (2009) extend the original CBD model by adding a cohort effect and a quadratic age effect to obtain the predictor:

$$\eta_{xt} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \left((x - \bar{x})^2 - \hat{\sigma}_x^2\right)\kappa_t^{(3)} + \gamma_{t-x}, \tag{10}$$

where $\hat{\sigma}_x^2$ is the average value of $(x - \bar{x})^2$. This model, usually referred to as model M7, is not identifiable as the transformation

$$\left(\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_{t-x}\right) \rightarrow \left(\kappa_t^{(1)} + \phi_1 + \phi_2(t - \bar{x}) + \phi_3\left((t - \bar{x})^2 + \hat{\sigma}_x^2\right), \kappa_t^{(2)} - \phi_2 - 2\phi_3(t - \bar{x}),\right.$$

$$\left.\kappa_t^{(3)} + \phi_3, \gamma_{t-x} - \phi_1 - \phi_2(t - x) - \phi_3(t - x)^2\right), \tag{11}$$

for real constants ϕ_1, ϕ_2 and ϕ_3, leaves the predictor unchanged. To identify the model Cairns *et al.* (2009) impose the *set of constraints*:

$$\sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c = 0, \qquad \sum_{c=t_1-x_k}^{t_n-x_1} c\gamma_c = 0, \qquad \sum_{c=t_1-x_k}^{t_n-x_1} c^2\gamma_c = 0,$$

which ensure that the cohort effect fluctuates around zero and has no discernible linear or quadratic trend. Following Haberman and Renshaw (2011, Appendix A), these constraints can be imposed by applying transformation (11) with constants ϕ_1, ϕ_2 and ϕ_3 obtained by regressing γ_{t-x} on $t - x$ and $(t - x)^2$, so that

$$\gamma_{t-x} = \phi_1 + \phi_2(t - x) + \phi_3(t - x)^2 + \epsilon_{t-x}, \quad \epsilon_{t-x} \sim N(0, \sigma^2) \quad \text{i.i.d..}$$

Cairns *et al.* (2009) also consider the simpler predictor structures

$$\eta_{xt} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \gamma_{t-x},$$

$$\eta_{xt} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + (x_c - x)\gamma_{t-x},$$

where x_c is a constant parameter to be estimated. These structures are typically referred to as models M6 and M8, respectively.

3.6. Plat model

Plat (2009) combines the CBD model with some features of the Lee-Carter model to produce a model that is suitable for full age ranges and captures the cohort effect. The proposed predictor structure assumes that there is a static age function, α_x, three age-period terms ($N = 3$) with pre-specified age-modulating parameters $\beta_x^{(1)} = 1$, $\beta_x^{(2)} = \bar{x} - x$, $\beta_x^{(3)} = (\bar{x} - x)^+ = \max(0, \bar{x} - x)$, and a cohort effect with pre-specified age-modulating parameters $\beta_x^{(0)} = 1$. Thus, the predictor is given by:

$$\eta_{xt} = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + (\bar{x} - x)^+\kappa_t^{(3)} + \gamma_{t-x}. \tag{12}$$

Plat (2009) targets the force of mortality μ_{xt} with the log *link* and estimates the parameters of the model by assuming a Poisson distribution of the deaths. The following parameter

transformations leave the predictor in (12) unchanged:

$$\left(\alpha_x, \kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_{t-x}\right) \rightarrow \left(\alpha_x + \phi_1 - \phi_2 x + \phi_3 x^2, \kappa_t^{(1)} + \phi_2 t + \phi_3(t^2 - 2\bar{x}t),\right.$$
$$\left. \kappa_t^{(2)} + 2\phi_3 t, \kappa_t^{(3)}, \gamma_{t-x} - \phi_1 - \phi_2(t-x) - \phi_3(t-x)^2\right) \quad (13)$$

$$\left(\alpha_x, \kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_{t-x}\right) \rightarrow \left(\alpha_x + c_1 + c_2(\bar{x} - x) + c_3(\bar{x} - x)^+,\right.$$
$$\left. \kappa_t^{(1)} - c_1, \kappa_t^{(2)} - c_2, \kappa_t^{(3)} - c_3, \gamma_{t-x}\right), \quad (14)$$

where c_1, c_2, c_3, ϕ_1, ϕ_2, and ϕ_3 are any real constants. The following *set of parameter constraints* can be imposed to ensure identifiability:

$$\sum_t \kappa_t^{(1)} = 0, \sum_t \kappa_t^{(2)} = 0, \sum_t \kappa_t^{(3)} = 0, \sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c = 0, \sum_{c=t_1-x_k}^{t_n-x_1} c\gamma_c = 0, \sum_{c=t_1-x_k}^{t_n-x_1} c^2\gamma_c = 0 \quad (15)$$

The first three constraints ensure that the period indexes are centered around zero, while the last three constraints ensure that the cohort effect fluctuates around zero and has no linear or quadratic trend. Following Haberman and Renshaw (2011, Appendix A), the constraints on the cohort effect can be imposed by applying transformation (13) with constants ϕ_1, ϕ_2, and ϕ_3 obtained by regressing γ_{t-x} on $t - x$ and $(t - x)^2$, so that

$$\gamma_{t-x} = \phi_1 + \phi_2(t - x) + \phi_3(t - x)^2 + \epsilon_{t-x}, \quad \epsilon_{t-x} \sim N(0, \sigma^2) \quad \text{i.i.d..} \quad (16)$$

The constraints on the period indexes can then be imposed by applying transformation (14) with

$$c_i = \frac{1}{n}\sum_t \kappa_t^{(i)}, \qquad i = 1, 2, 3. \quad (17)$$

In the cases where only older ages are of interest, Plat (2009) suggests to drop the third period term from predictor (12):[4]

$$\eta_{xt} = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + \gamma_{t-x}. \quad (18)$$

We note that this reduced Plat model has the same identifiability issues as the complete model with the omission of the transformations and constraints involving $\kappa_t^{(3)}$ and c_3.

---

[4]Note that the reduced Plat model is essentially the M6 model of Cairns *et al.* (2009) with an added static age term α_x.

4. GAPC stochastic mortality models with StMoMo

The **StMoMo** package provides an R implementation of the GAPC family of stochastic mortality models using the standard S3 object-oriented system. **StMoMo** can be installed with the code:

```
R> install.packages("StMoMo")
```

The package is loaded within R as follows:

```
R> library("StMoMo")
```

In the package **StMoMo**, GAPC stochastic mortality models are constructed using the StMoMo function. The synopsis of this functions is:

```
 StMoMo(link = c("log", "logit"), staticAgeFun = TRUE, periodAgeFun = "NP",
   cohortAgeFun = NULL, constFun = function(ax, bx, kt, b0x, gc, wxt, ages)
   list(ax = ax, bx = bx, kt = kt, b0x = b0x, gc = gc))
```

The StMoMo function takes as input information on the *link function* (and the associated distributional assumption), the *predictor structure*, and the *set of parameter constraints* to create an object of the type "StMoMo" representing a GAPC mortality model:

- The argument link defines the *link function* and *the random component* associated with the mortality model. Setting link = "log" assumes that deaths follow a Poisson distribution and uses a log link targeting the force of mortality μ_{xt}, while setting link = "logit" assumes that deaths follow a Binomial distribution and uses a logit link targeting one-year death probabilities q_{xt}.

- The *predictor* of the model is defined via the arguments staticAgeFun, periodAgeFun and cohortAgeFun. Argument staticAgeFun is a logical variable indicating whether the model has a static age function α_x or not. Argument periodAgeFun is a list of length N containing the definitions of the period age-modulating parameters $\beta_x^{(i)}$, $i = 1, \ldots, N$, with each entry being either "NP" for non-parametric age terms, "1" for $\beta_x^{(i)} = 1$, or a predefined parametric function of age.[5] Argument cohortAgeFun defines the cohort age modulating parameter $\beta_x^{(0)}$ and can take values "NP" for non-parametric age terms, "1" for $\beta_x^{(0)} = 1$, a predefined parametric function of age, or NULL if the model does not have a cohort effect.

- The *set of parameter constraints* are defined via the argument constFun which is a user-defined implementation of the *constraint function v* mapping an arbitrary vector of parameters to a vector of transformed parameters satisfying the model constraints.

We note that due to limitations of the R functions used for fitting "StMoMo" objects to data (see Section 5), the current version of **StMoMo** does not support models combining parametric and non-parametric age-modulating functions, $\beta_x^{(i)}$, $0 = 1, \ldots, N$. However, such models

---

[5]Note that we can define a model with no age-period terms ($N = 0$) by making periodAgeFun = NULL.

Model	Predictor
LC	$\eta_{xt} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)}$
CBD	$\eta_{xt} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)}$
APC	$\eta_{xt} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x}$
RH	$\eta_{xt} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} + \gamma_{t-x}$
M7	$\eta_{xt} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \left((x - \bar{x})^2 - \hat{\sigma}_x^2\right) \kappa_t^{(3)} + \gamma_{t-x}$
PLAT	$\eta_{xt} = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + \gamma_{t-x}$

Table 1: Model structures considered in this paper.

are not typically considered and the majority of models proposed in the literature are either extensions of the Lee-Carter model with all age-modulating terms being non-parametric or extensions of the CBD model with all age-modulating terms being parametric.[6]

In order to illustrate the creation of particular GAPC mortality models and other capabilities of **StMoMo**, in the rest of this paper we will focus on the models summarized in Table 1. From now on, LC stands for the Lee-Carter model; CBD for the original Cairns-Blake-Dowd model; APC for the age-period-cohort model; RH for the cohort extension of the Lee-Carter model defined in Equation 7 and proposed by Renshaw and Haberman (2006); M7 for the quadratic CBD model defined in Equation 10; and PLAT for the reduced Plat model defined previously in Equation 18. For the sake of comparability, in all cases we will assume a Binomial distribution of deaths and use the logit function to link q_{xt} to the predictor structure η_{xt}.

Below, we show how to define each of the models in Table 1 using the package **StMoMo**.

Lee-Carter model

The LC model under a Binomial setting can be defined using the following code:

```
R> constLC <- function(ax, bx, kt, b0x, gc, wxt, ages){
+    c1 <- mean(kt[1, ], na.rm = TRUE)
+    c2 <- sum(bx[, 1], na.rm = TRUE)
+    list(ax = ax + c1 * bx, bx = bx / c2, kt =  c2 * (kt - c1))
+ }
R> LC <- StMoMo(link = "logit", staticAgeFun = TRUE, periodAgeFun = "NP",
+    constFun = constLC)
```

Recalling Section 3.1, we note that the constraint function `constLC` is the R implementation of transformation (2) with constants c_1 and c_2 calculated using Equation 4 to impose the

---

[6]For instance, a model with predictor structure $\eta_{xt} = \alpha_x + (x - \bar{x})\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(1)}$, corresponding to `StMoMo(staticAgeFun = TRUE, periodAgeFun = c(f1, "NP"))`, with `f1 <- function(x, ages) x - mean(ages)`, is not supported.

constraints defined in Equation 3. The **StMoMo** package also contains the function `lc` to facilitate the definition of Lee-Carter models. Hence, we could define the LC model using the much simpler predefined command:

```
R> LC <- lc(link = "logit")
```

CBD model

To define the CBD model we use the following commands:

```
R> f2 <- function(x, ages) x - mean(ages)
R> CBD <- StMoMo(link = "logit", staticAgeFun = FALSE,
+    periodAgeFun = c("1", f2))
```

Here, we note that function `f2` defines the second age-modulating parameter $\beta_x^{(2)} = x - \bar{x}$ and that a `constFUN` argument need not be provided since the CBD model does not have identifiability issues. Alternatively, we can define the CBD model using the predefined function `cbd`:

```
R> CBD <- cbd()
```

APC model, RH model and M7 model

The APC, RH and M7 models could be defined by implementing explicitly the discussions in Sections 3.2, 3.3 and 3.5. However, **StMoMo** includes predefined functions `apc`, `rh`, `m7` that facilitate the definition of the APC model, the RH model and model M7, respectively.[7] Thus, these models are defined with the code:

```
R> RH <- rh(link = "logit",  cohortAgeFun = "1")
R> APC <- apc(link = "logit")
R> M7 <- m7()
```

PLAT model

The package **StMoMo** does not include a predefined function for the Plat model. Nevertheless, recalling Section 3.6, we can define the reduced Plat model using the code:

```
R> f2 <- function(x, ages) mean(ages) - x
R> constPlat <- function(ax, bx, kt, b0x, gc, wxt, ages){
+    nYears <- dim(wxt)[2]
+    x <- ages
+    t <- 1:nYears
+    c <- (1 - tail(ages, 1)):(nYears - ages[1])
+    xbar <- mean(x)
```

---

[7]The **StMoMo** package also includes functions `m6` and `m8` for defining models M6 and M8. We also note that the Renshaw and Haberman (2006) cohort extension of the Lee-Carter in Equation 5 can be defined using the function `rh` with argument `cohortAgeFun = "NP"`.

```
+     phiReg <- lm(gc ~ 1 + c + I(c ^ 2), na.action = na.omit)
+     phi <- coef(phiReg)
+     gc <- gc - phi[1] - phi[2] * c - phi[3] * c ^ 2
+     kt[2, ] <- kt[2, ] + 2 * phi[3] * t
+     kt[1, ] <- kt[1, ] + phi[2] * t + phi[3] * (t ^ 2 - 2 * xbar * t)
+     ax <- ax + phi[1] - phi[2] * x + phi[3] * x ^ 2
+     ci <- rowMeans(kt, na.rm = TRUE)
+     ax <- ax + ci[1] + ci[2] * (xbar - x)
+     kt[1, ] <- kt[1, ] - ci[1]
+     kt[2, ] <- kt[2, ] - ci[2]
+     list(ax = ax, bx = bx, kt = kt, b0x = b0x, gc = gc)
+   }
R> PLAT <- StMoMo(link = "logit", staticAgeFun = TRUE,
+     periodAgeFun = c("1", f2), cohortAgeFun = "1", constFun = constPlat)
```

We note that the constraint function `constPlat` is the R implementation of transformations (13) and (14) omitting the terms involving $\kappa^{(3)}$ and c_3, and with constants ϕ_1, ϕ_2, ϕ_3 obtained via the linear regression defined in (16) and constants c_1 and c_2 as in Equation 17. Function `constPlat` imposes the constraints in Equation 15.

5. Model fitting

Parameter estimates of GAPC stochastic mortality models can be obtained by maximizing the model log-likelihood, which is given by

$$\mathcal{L}(d_{xt}, \hat{d}_{xt}) = \sum_x \sum_t \omega_{xt} \left\{ d_{xt} \log \hat{d}_{xt} - \hat{d}_{xt} - \log d_{xt}! \right\}$$

in the case of a Poisson distribution of deaths, and by

$$\mathcal{L}(d_{xt}, \hat{d}_{xt}) = \sum_x \sum_t \omega_{xt} \left\{ d_{xt} \log \left(\frac{\hat{d}_{xt}}{E_{xt}^0} \right) + (E_{xt}^0 - d_{xt}) \log \left(\frac{E_{xt}^0 - \hat{d}_{xt}}{E_{xt}^0} \right) + \log \binom{E_{xt}^0}{d_{xt}} \right\}$$

in the case of a Binomial distribution of deaths. In both cases, ω_{xt} are weights taking the value 0 if a particular (x, t) data cell is omitted or 1 if the cell is included, and

$$\hat{d}_{xt} = E_{xt} \, g^{-1} \left(\alpha_x + \sum_{i=1}^N \beta_x^{(i)} \kappa_t^{(i)} + \beta_x^{(0)} \gamma_{t-x} \right)$$

is the expected number of deaths predicted by the model, with g^{-1} denoting the inverse of the *link function g*.

In the mortality literature, maximization of the log-likelihood is typically performed using the Newton-Raphson iterative procedure tailored for each model (see, e.g., Brouhns *et al.* (2002), Renshaw and Haberman (2006) and Cairns *et al.* (2009)). This is in fact the approach implemented in the packages **ilc** and **LifeMetrics**. Nonetheless, as discussed extensively by

Currie (2016), many stochastic mortality models are examples of generalized linear models or generalized non-linear model, which facilitates their fitting using standard statistical software.[8] Currie (2016) exemplifies this fact by fitting several stochastic mortality models in R using the standard function `glm` or the function `gnm` of the package **gnm** (Turner and Firth 2015).[9]

StMoMo provides the generic function `fit` for estimating the parameters of GAPC mortality models. In line with the remarks of Currie (2016), the corresponding S3 method for objects of the class `"StMoMo"` heavily relies on the function `gnm` of the package **gnm** to estimate the parameters of a GAPC model.Internally, this is accomplished by constructing the equivalent **gnm** formulation of the GAPC mortality model.[10] For instance, the **gnm** formula of the Binomial `LC` model created before is

```
R> LC$gnmFormula
```

```
[1] "D/E ~ -1 + offset(o) + factor(x) + Mult(factor(x), factor(t), inst = 1)"
```

while the **gnm** formula of the Binomial `CBD` model defined before is

```
R> CBD$gnmFormula
```

```
[1] "D/E ~ -1 + offset(o) + factor(t) + B2:factor(t)"
```

We now illustrate the usage of the function `fit` of the package **StMoMo** by fitting the six models defined before to England and Wales mortality data. Function `fit` expects the user to provide a list of class `"StMoMoData"` containing deaths and exposures in a matrix format with ages on the rows and calendar years on the columns. Such a type of list can readily be created using the `StMoMoData` function as exemplified in Section 8. For illustration purposes, the object `EWMaleData`, included in the package **StMoMo**, contains deaths counts (`EWMaleData$Dxt`) and central exposures (`EWMaleData$Ext`) for England and Wales males for the period 1961-2011 and for ages 0-100 obtained from the Human Mortality Database (2014).

```
R> EWMaleData
```

```
Mortality data for England and Wales
    Series:  male
    Years: 1961 - 2011
    Ages:  0 - 100
    Exposure:  central
```

---

[8] Haberman and Renshaw (2011) have also noticed this fact and profit from GLM facilities in standard statistical packages when fitting CBD type models.

[9] Debón *et al.* (2010) also discuss the use of the package **gnm** for fitting Lee-Carter type models.

[10] We note that when all the $\beta_x^{(i)}$ are parametric functions of age, the model is a GLM and therefore **gnm** by default resorts to the `glm` function of R when fitting the parameters of the model.

However, in our examples we concentrate on ages 55 to 89 as the CBD model and the M7 model have been particularly designed to fit higher ages. Additionally, since some models include cohort effects and in agreement with the usual practice (see e.g Cairns *et al.* (2009) and Haberman and Renshaw (2011)), we exclude (by setting $\omega_{xt} = 0$) all cohorts that have fewer than three observations. Missing values for either death counts or exposures are automatically zero-weighted.

Models LC, APC, CBD, M7 and PLAT can be fitted to England and Wales male mortality data for ages 55 to 89 using the code:

```
R> EWMaleIniData <- central2initial(EWMaleData)
R> ages.fit <- 55:89
R> wxt <- genWeightMat(ages = ages.fit, years = EWMaleIniData$years,
+    clip = 3)
R> LCfit <- fit(LC, data = EWMaleIniData, ages.fit = ages.fit, wxt = wxt)
R> APCfit <- fit(APC, data = EWMaleIniData, ages.fit = ages.fit, wxt = wxt)
R> CBDfit <- fit(CBD, data = EWMaleIniData, ages.fit = ages.fit, wxt = wxt)
R> M7fit <- fit(M7, data = EWMaleIniData, ages.fit = ages.fit, wxt = wxt)
R> PLATfit <- fit(PLAT, data = EWMaleIniData, ages.fit = ages.fit, wxt = wxt)
```

From this code we note the following:

- In order to match the logit-Binomial setting used before in the definition of the mortality models, initial exposures are approximated by transforming the available central exposures. This is accomplished using the utility function `central2initial` of package **StMoMo**.

- The first and last three cohorts years are excluded from the fitting via the argument `wxt`. The appropriate 0-1 weighting matrix, `wxt`, is constructed using the utility function `genWeightMat` of package **StMoMo**.

The fitting of the RH model requires some care as it is well known that fitting cohort extensions of the Lee-Carter models is problematic (Hunt and Villegas 2015). In particular, Currie (2016) has encountered convergence issues when using package **gnm** to fit the RH model. As a possible way to circumvent these issues, Currie (2016) suggests the use of appropriate starting values when fitting model RH. This can be achieved in the function `fit` via input arguments `start.ax`, `start.bx`, `start.kt`, `start.b0x`, and `start.gc`. Using the parameters of the Lee-Carter model as starting values, model RH can be fitted with the code:[11]

```
R> RHfit <- fit(RH, data = EWMaleIniData,  ages.fit = ages.fit, wxt = wxt,
+    start.ax = LCfit$ax, start.bx = LCfit$bx, start.kt = LCfit$kt)
```

The output from the function `fit` is an object of the class `"fitStMoMo"` including, among other things, the following information:

---

[11]**StMoMo** also implements the fitting method suggested by Hunt and Villegas (2015) to solve the convergence issues of the RH model. This can be acomplished by setting argument `approxConst = TRUE` when defining the RH model, that is, `rh(approxConst=TRUE)`.
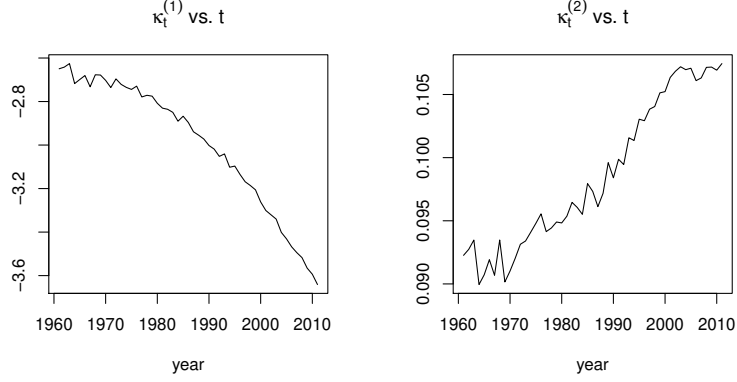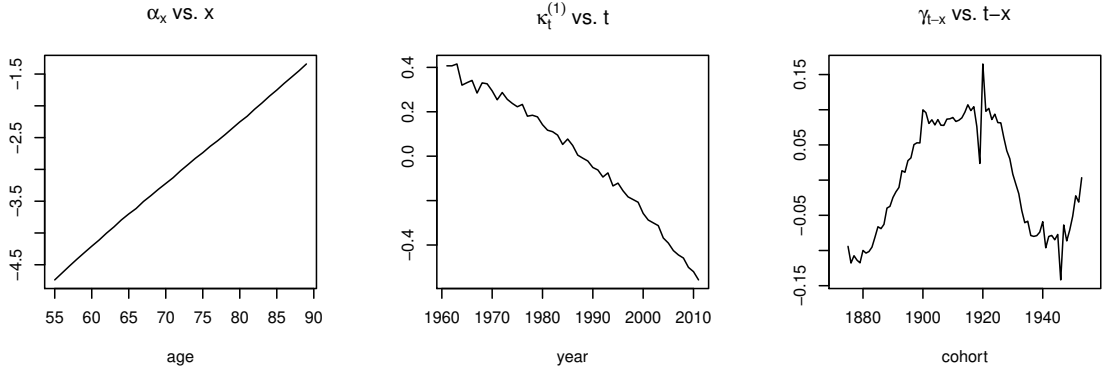
Figure 1: Parameters for the Lee-Carter (LC) model fitted to the England and Wales male population for ages 55-89 and the period 1961-2011.

- `model`: the `"StMoMo"` object defining the underlying GAPC stochastic mortality model;

- `ax`, `bx`, `kt`, `b0x`, `gc`: the estimated parameters;

- `loglik`: the log-likelihood of the model;

- `deviance`: the model deviance;

- `nobs`: the number of observations in the data;

- `npar`: the effective number of parameters of the model;

- `fittingModel`: the output of the `gnm` call used to fit the model.

There are `print`, `plot`, `fitted`, and `residuals` methods for the `"fitStMoMo"` class. For instance, Figures 1, 2 and 3 depicting the fitted parameters of the LC model, the CBD model and the APC model, respectively, were produced with the code:

```
R> plot(LCfit, nCol = 3)
R> plot(CBDfit, parametricbx = FALSE)
R> plot(APCfit, parametricbx = FALSE, nCol = 3)
```

6. Goodness-of-fit analysis

The goodness-of-fit of mortality models is typically analyzed by inspecting the residuals of the fitted model. Regular patterns in the residuals indicate the inability of the model to describe all the features of the data appropriately. With a Poisson or Binomial random component, it is appropriate to look at the scaled deviance residuals defined as:

$$r_{xt} = \text{sign}(d_{xt} - \hat{d}_{xt})\sqrt{\frac{\text{dev}(x,t)}{\hat{\phi}}}, \qquad \hat{\phi} = \frac{D(d_{xt}, \hat{d}_{xt})}{K - \nu},$$

Figure 2: Parameters for the CBD model fitted to the England and Wales male population for ages 55-89 and the period 1961-2011.

Figure 3: Parameters for the APC model fitted to the England and Wales male population for ages 55-89 and the period 1961-2011.

where

$$\text{dev}(x,t) = 2\left[d_{xt}\log\left(\frac{d_{xt}}{\hat{d}_{xt}}\right) - (d_{xt} - \hat{d}_{xt})\right]$$

for a Poisson random component, or

$$\text{dev}(x,t) = 2\left[d_{xt}\log\left(\frac{d_{xt}}{\hat{d}_{xt}}\right) + (E_{xt}^0 - d_{xt})\log\left(\frac{E_{xt}^0 - d_{xt}}{E_{xt}^0 - \hat{d}_{xt}}\right)\right]$$

for a Binomial random component. Further,

$$D(d_{xt}, \hat{d}_{xt}) = \sum_x \sum_t \omega_{xt}\text{dev}(x,t)$$

is the total deviance of the model, $K = \sum_x \sum_t \omega_{xt}$ is the number of observations in the data and ν is the effective number of parameters in the model.

In **StMoMo** standardized deviance residuals can be obtained with the generic function `residuals` applied to a fitted stochastic mortality model of the class `"fitStMoMo"`. For example, to obtain the residuals of the LC model and of the CBD model fitted before we use the commands:

```
R> LCres <- residuals(LCfit)
R> CBDres <- residuals(CBDfit)
```

Graphs of these residuals can be produced using the generic function `plot`. This function supports, via the argument `type`, three types of plots:

- scatter plots of residuals by age, period and cohort such as those extensively used in Haberman and Renshaw (2011);

- black and white sign-plots of the residuals such as those used in Cairns *et al.* (2009) and Lovász (2011); and

- color maps (heat-maps) of the residuals.

Figure 4 presents heat-maps of the deviance residuals for the six models fitted to the England and Wales male mortality experience. These charts were produced using function `plot` with option `type = "colourmap"`. For instance, Figure 4a was obtained with the code:

```
R> plot(LCres, type = "colourmap", reslim = c(-3.5, 3.5))
```

From Figure 4 we see that models LC, CBD and APC display strong residual patterns while the residuals of models RH, M7 and PLAT look reasonably random. The APC model shows a strong clustering of residuals due to its inability to allow for varying improvement rates with age. The LC and CBD models, which do not incorporate a cohort effect, show very marked diagonals patterns indicating the inability of these models to capture the well-known cohort effect observed in the England and Wales population (Willets 2004). The issues with the fit of the LC and CBD models become more evident when looking at scatter plots of the residuals by age, period and cohort. Such plots for the LC model (Figure 5a) and the CBD model (Figure 5b) can be produced using argument `type = "scatter"` of the function `plot` via the commands:

```
R> plot(LCres, type = "scatter", reslim = c(-3.5, 3.5))
R> plot(CBDres, type = "scatter", reslim = c(-3.5, 3.5))
```

The right panels in Figure 5 clearly show that the LC and CBD models are unable to capture the cohort effect. In addition, the left panel in Figure 5b reveals some strong patterns by age, reflecting the lack of a quadratic age term in the CBD which may be necessary to capture the commonly observed curvature of the mortality rates in a logit scale.

When evaluating the goodness-of-fit of different models, it is generally anticipated that models with more parameters provide a better fit to the data. To rule out the possibility that the better fit observed in a model is the result of over-parametrization and compare the relative performance of several models, it has become common in the mortality literature to use information criteria which modify the maximum likelihood criterion by penalizing models
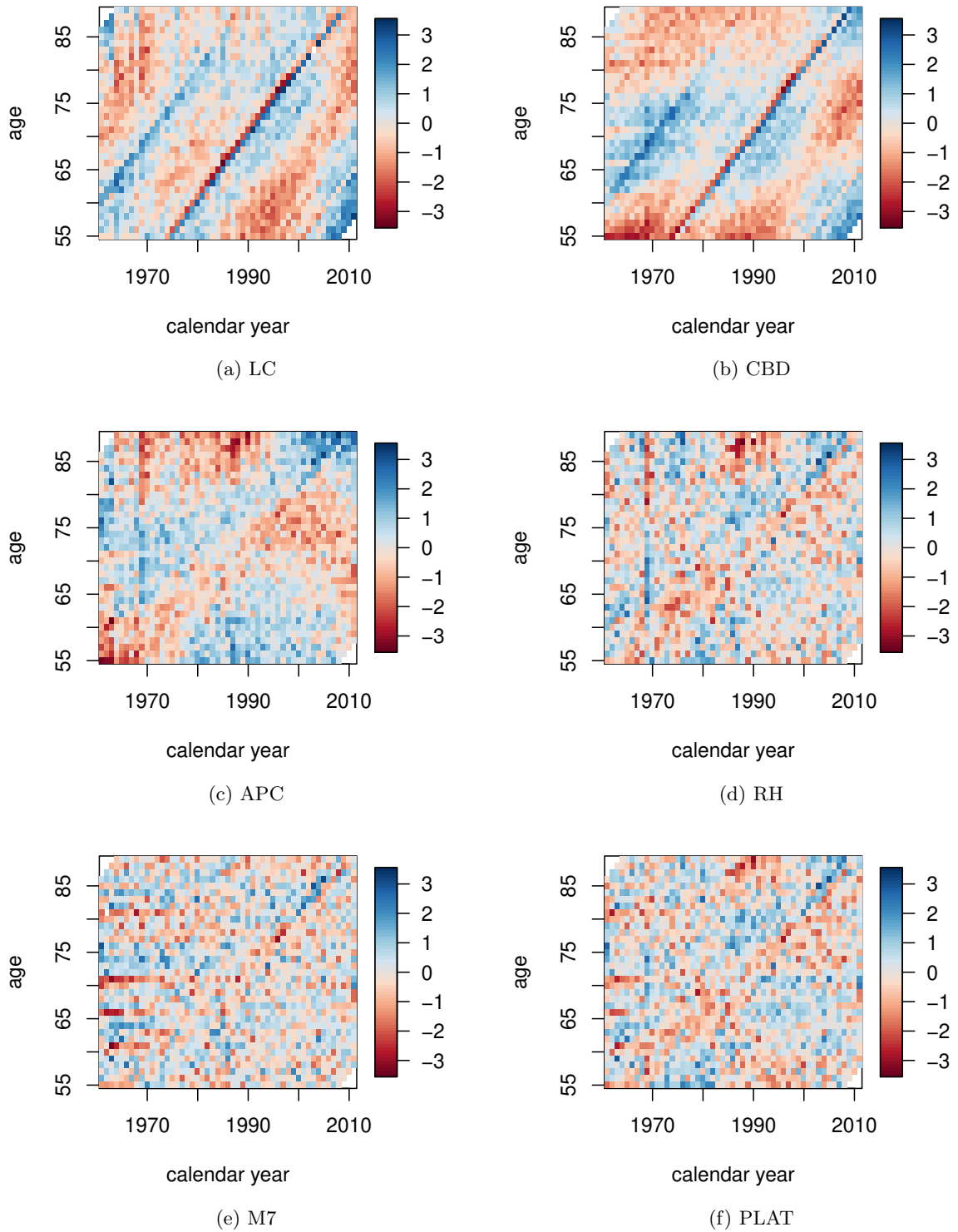
Figure 4: Heat-maps of deviance residuals for different model fitted to the England and Wales males population for ages 55-89 and the period 1961-2011.
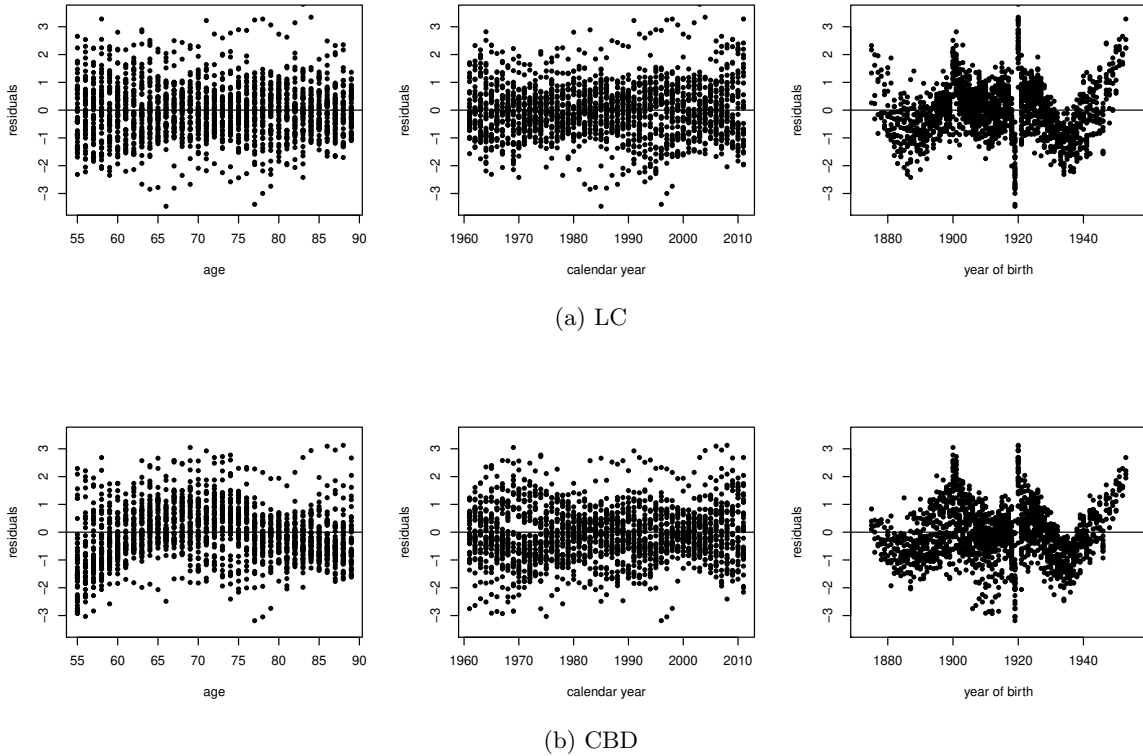
(a) LC

(b) CBD

Figure 5: Scatter plots of deviance residuals for models LC and CBD fitted to the England and Wales males population for ages 55-89 and the period 1961-2011.

with more parameters.[12] Two of these criteria are the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC), defined as $AIC = 2\nu - 2\mathcal{L}$ and $BIC = \nu \log K - 2\mathcal{L}$, respectively, with a lower value of AIC and BIC being preferable. In R these information criteria can be computed using the generic functions `AIC` and `BIC`. For example, we can get the AIC and BIC of the CBD model as follows:

```
R> AIC(CBDfit)
```

```
[1] 34697.82
```

```
R> BIC(CBDfit)
```

```
[1] 35256.83
```

Table 2 presents AIC and BIC values for the six models fitted to the England and Wales male data. We note that both criteria lead to the same ranking of models with M7, PLAT, and RH being the best performing models. Overall, these results are consistent with the existing literature comparing single population models, where the Renshaw-Haberman extension of the

---

[12]For examples of such analysis see Cairns *et al.* (2009, Section 6.1.1), Haberman and Renshaw (2011, Section 3.3), and Lovász (2011, Section 4.1).

Lee-Carter model and the M7 model have been identified as good candidates for modeling mortality in the England and Wales population (Cairns *et al.* 2009; Haberman and Renshaw 2011).

	LC	CBD	APC	RH	M7	PLAT
Number of parameters	119	102	162	197	229	211
AIC	29866(5)	34698(6)	24469(4)	21779(3)	21406(1)	21624(2)
BIC	30518(5)	35257(6)	25357(4)	22859(3)	22661(1)	22780(2)

Table 2: Number of parameters, AIC and BIC values (with their respective rankings in brackets) for different model fitted to the England and Wales males population for ages 55-89 and the period 1961-2011.

7. Forecasting and simulation with stochastic mortality models

In the family of GAPC stochastic mortality models the dynamics of mortality are driven by the period indexes $\kappa_t^{(i)}$, $i = 1, \ldots, N$, and the cohort index γ_{t-x}. Therefore, the forecasting and simulation of mortality rates requires the modeling of these indexes using time series techniques.

For the period indexes we consider two alternative modelling approaches. A first possibility is to use the standard approach in the actuarial literature (Cairns *et al.* 2006, 2011; Haberman and Renshaw 2011; Lovász 2011) and assume that the period indexes follow a multivariate random walk with drift. That is,

$$\boldsymbol{\kappa}_t = \boldsymbol{\delta} + \boldsymbol{\kappa}_{t-1} + \boldsymbol{\xi}_t^\kappa, \qquad \boldsymbol{\kappa}_t = \begin{pmatrix} \kappa_t^{(1)} \\ \vdots \\ \kappa_t^{(N)} \end{pmatrix}, \qquad \boldsymbol{\xi}_t^\kappa \sim N(\mathbf{0}, \Sigma), \tag{19}$$

where $\boldsymbol{\delta}$ is an N-dimension vector of drift parameters and Σ is the $N \times N$ variance-covariance matrix of the multivariate white noise $\boldsymbol{\xi}_t^\kappa$.

A second alternative is to assume that the individual period indexes, $\kappa_t^{(i)}$, $i = 1, \ldots, N$, follow a general univariate ARIMA model. Under this approach, the i-th period index, $\kappa_t^{(i)}$, is assumed to follow an ARIMA(p_i, q_i, d_i) with drift, so that

$$\Delta^{d_i} \kappa_t^{(i)} = \delta_0^{(i)} + \phi_1^{(i)} \Delta^{d_i} \kappa_{t-1}^{(i)} + \cdots + \phi_{p_i}^{(i)} \Delta^{d_i} \kappa_{t-p_i}^{(i)} + \xi_t^{(i)} + \delta_1^{(i)} \xi_{t-1}^{(i)} + \cdots + \delta_{q_i}^{(i)} \xi_{t-q_i}^{(i)}, \tag{20}$$

where Δ is the difference operator, $\delta_0^{(i)}$ is the drift parameter, $\phi_1^{(i)}, \ldots, \phi_{p_i}^{(i)}$ are the autoregressive coefficients with $\phi_{p_i} \neq 0$, $\delta_1^{(i)}, \ldots, \delta_{q_i}^{(i)}$ are the moving average coefficients with $\delta_{q_i}^{(i)} \neq 0$ and $\xi_t^{(i)}$ is a Gaussian white noise process with variance $\sigma_\xi^{(i)}$.

As pointed out by Currie (2016), the main challenge when forecasting stochastic mortality models is specifying the dynamics of the cohort effect. To have a simple starting point, we

Mortality model	Model for γ_{t-x}
APC	$\mathrm{ARIMA}(1,1,0)$ with drift
RH	$\mathrm{ARIMA}(1,1,0)$ with drift
M7	$\mathrm{ARIMA}(2,0,0)$ with non-zero intercept
PLAT	$\mathrm{ARIMA}(2,0,0)$ with non-zero intercept

Table 3: ARIMA models for the cohort effect for models APC, RH, M7 and PLAT.

follow previous studies (Renshaw and Haberman 2006; Cairns *et al.* 2011; Lovász 2011) and assume that the cohort index, γ_{t-x}, follows a univariate ARIMA process which is independent of the period index, $\boldsymbol{\kappa}_t$. In general, we assume that $\gamma_c \equiv \gamma_{t-x}$ follows an $\mathrm{ARIMA}(p,q,d)$ with drift, so that

$$\Delta^d \gamma_c = \delta_0 + \phi_1 \Delta^d \gamma_{c-1} + \cdots + \phi_p \Delta^d \gamma_{c-p} + \epsilon_c + \delta_1 \epsilon_{c-1} + \cdots + \delta_q \epsilon_{c-q}, \qquad (21)$$

where ϵ_c is a Gaussian white noise process with variance σ_ϵ.

The time series models in Equations 19, 20 and 21 can be used to obtain projected (simulated) values of the period index $\dot{\boldsymbol{\kappa}}_{t_n+s} := \left(\dot{\kappa}_{t_n+s}^{(1)}, \ldots, \dot{\kappa}_{t_n+s}^{(N)} \right)'$ and cohort index $\dot{\gamma}_{t_n+s-x_1}$, $s = 1, \ldots, h$, to derive forecasted (simulated) values of the predictor

$$\dot{\eta}_{x,t_n+s} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \dot{\kappa}_{t_n+s}^{(i)} + \beta_x^{(0)} \dot{\gamma}_{t_n+s-x},$$

which can in turn be used to obtain forecasted (simulated) age-specific central mortality rates, $\dot{\mu}_{x,t_n+s}$ or age-specific one-year death probabilities, $\dot{q}_{x,t_n+s}$.

In the package **StMoMo** the forecasting of GAPC stochastic mortality models is implemented via the generic method `forecast`. This function estimates and forecasts the multivariate random walk with drift in Equation 19 using the approach described in Haberman and Renshaw (2011, Appendix B)[13] and uses function `Arima` of package **forecast** (Hyndman and Khandakar 2008; Hyndman 2017) to estimate and forecast the ARIMA processes of Equation 20 and Equation 21. For instance, if we assume that the period indexes follow a multivariate random walk with drift (the default in **StMoMo**) and that the cohort indexes of the APC, RH, M7, and PLAT model follow the ARIMA processes specified in Table 3, 50-year ahead ($h = 50$) central projections of the period indexes, cohort index, and one-year death probabilities for the England and Wales mortality experience can be obtain with the code:

```
R> LCfor <- forecast(LCfit, h = 50)
R> CBDfor <- forecast(CBDfit, h = 50)
R> APCfor <- forecast(APCfit, h = 50, gc.order = c(1, 1, 0))
```

---

[13]This is implemented in the function `mrwd` of the package **StMoMo**.

```
R> RHfor <- forecast(RHfit, h = 50, gc.order = c(1, 1, 0))
R> M7for <- forecast(M7fit, h = 50, gc.order = c(2, 0, 0))
R> PLATfor <- forecast(PLATfit, h = 50, gc.order = c(2, 0, 0))
```

Alternatively, we could assume that the period indexes follow independent univariate ARIMA models. This is achieved in function `forecast` by setting the argument `method = "iarima"`. For example, projections of the Lee-Carter model under the assumption that $\kappa_t^{(1)}$ follows an ARIMA(1,1,2) with drift are produced with the code:

```
R> LCforArima <- forecast(LCfit, h = 50, kt.method = "iarima",
+     kt.order = c(1, 1, 2))
```

By setting `kt.order = NULL` the function `auto.arima` from package **forecast** selects the best ARIMA process for each period index.

The output from the function `forecast` is an object of the class `"forStMoMo"` including, among other things, the following information:

- `rates`: a matrix with the central projection of the mortality rates, $\dot{\mu}_{x,t_n+s}$ or $\dot{q}_{x,t_n+s}$, $s = 1, \ldots, h$;

- `kt.f`: a list containing information on the multivariate random walk with drift fitted to the period index $\boldsymbol{\kappa}_t$; and

- `gc.f`: a list containing information on the ARIMA model fitted to the cohort index γ_{t-x}.

There are `print` and `plot` methods for the `"forStMoMo"` class. For instance, plots of the forecast of the period indexes of models LC, RH, M7 and PLAT (see Figures 6 and 7) can be produced using the code:

```
R> plot(LCfor, only.kt = TRUE)
R> plot(LCforArima, only.kt = TRUE)
R> plot(RHfor, only.kt = TRUE)
R> plot(M7for, only.kt = TRUE, nCol = 3)
R> plot(PLATfor, only.kt = TRUE)
```

Similarly, plots of the forecast of the cohort indexes of the APC, RH, M7 and PLAT models (see Figure 8) can be obtained with the commands:
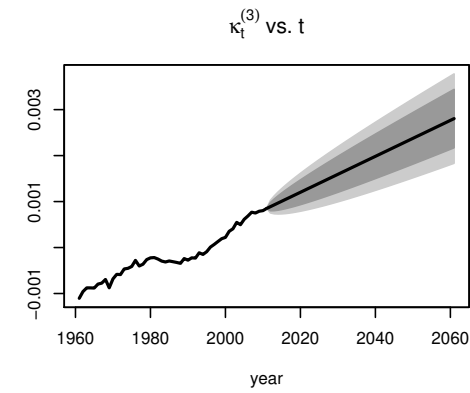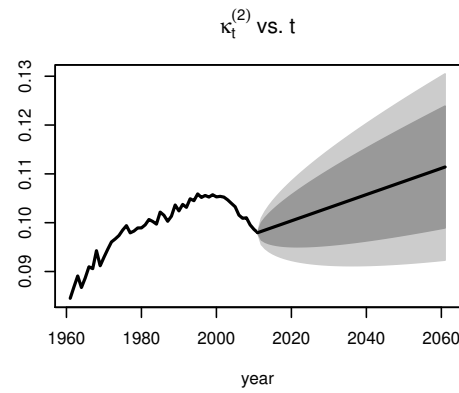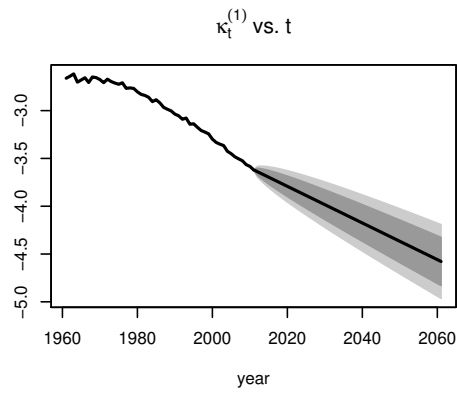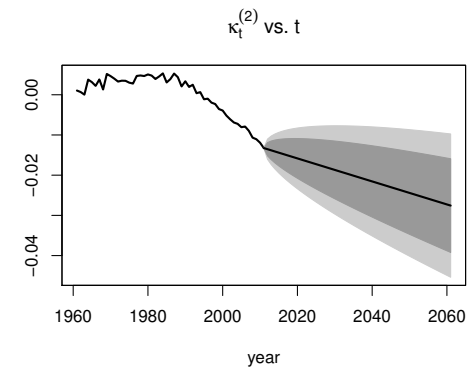
```
R> plot(APCfor, only.gc = TRUE)
R> plot(RHfor, only.gc = TRUE)
R> plot(M7for, only.gc = TRUE)
R> plot(PLATfor, only.gc = TRUE)
```
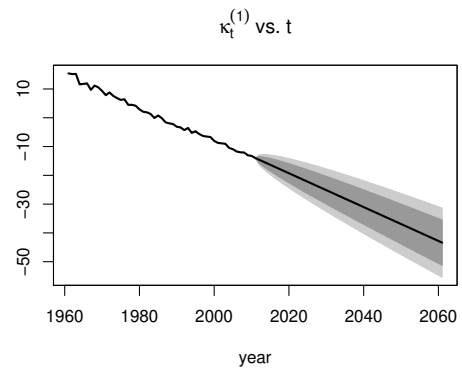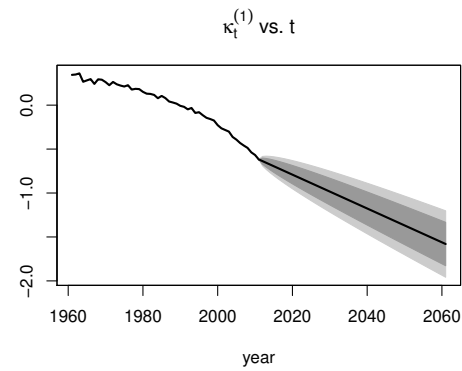
Figure 6: Forecast of the period indexes of the RH, M7 and PLAT models applied to the England and Wales males population for ages 55-89 and the period 1961-2011. Dashed lines represent central forecast and dotted lines represent 95% prediction intervals.
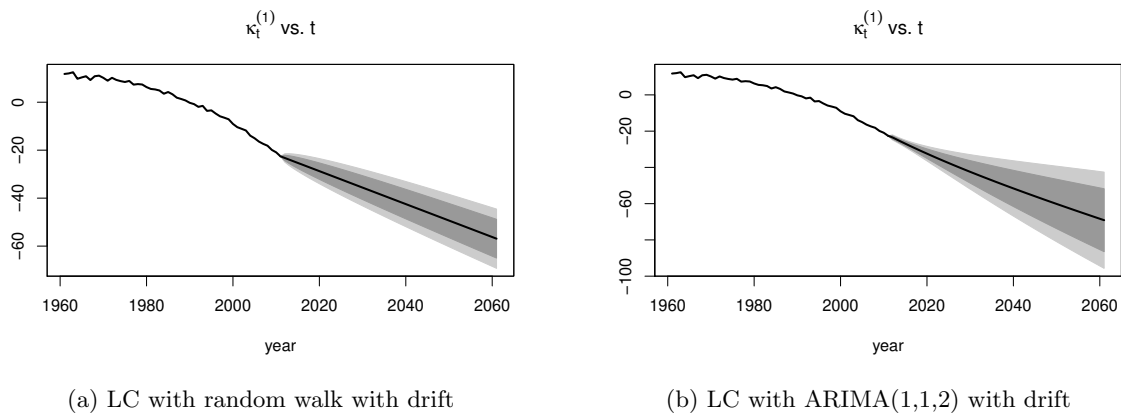
(a) LC with random walk with drift

(b) LC with ARIMA(1,1,2) with drift

Figure 7: Forecast of the period indexes of the LC model with random walk with drift and ARIMA(1,1,2) with drift applied to the England and Wales males population for ages 55-89 and the period 1961-2011. Dashed lines represent central forecast and dotted lines represent 95% prediction intervals.

The package **StMoMo** also provides the function `simulate` for simulating trajectories from GAPC stochastic mortality models. In order to simulate the period index, κ_t, in **StMoMo** we implement a multivariate adaptation of Algorithm 2 in Haberman and Renshaw (2009) without provision for parameter error,[14] while to simulate the cohort index, γ_{t-x},[15] function `simulate` uses the equivalent S3 method for objects of class `"Arima"` provided by the package **forecast**. For example, the code below produces 500 simulated trajectories for the next 50 years of the six stochastic mortality models fitted previously to the England and Wales male mortality experience:

```
R> set.seed(1234)
R> nsim <- 500
R> LCsim <- simulate(LCfit, nsim = nsim, h = 50)
R> CBDsim <- simulate(CBDfit, nsim = nsim, h = 50)
R> APCsim <- simulate(APCfit, nsim = nsim, h = 50, gc.order = c(1, 1, 0))
R> RHsim <- simulate(RHfit, nsim = nsim, h = 50, gc.order = c(1, 1, 0))
R> M7sim <- simulate(M7fit, nsim = nsim, h = 50, gc.order = c(2, 0, 0))
R> PLATsim <- simulate(PLATfit, nsim = nsim, h = 50, gc.order = c(2, 0, 0))
```

The output from the function `simulate` is an object of the class `"simStMoMo"` including, among other things, the following information:

- `rates`: a three dimensional array with the future simulated mortality rates;

- `kt.s`: a list containing information on the simulated paths of the period index κ_t; and

- `gc.s`: a list containing information on the simulated paths of the cohort index γ_{t-x}.

---

[14]We note that Algorithm 2 in Haberman and Renshaw (2009) is itself an adaptation of the prediction interval approach of Cairns *et al.* (2006).

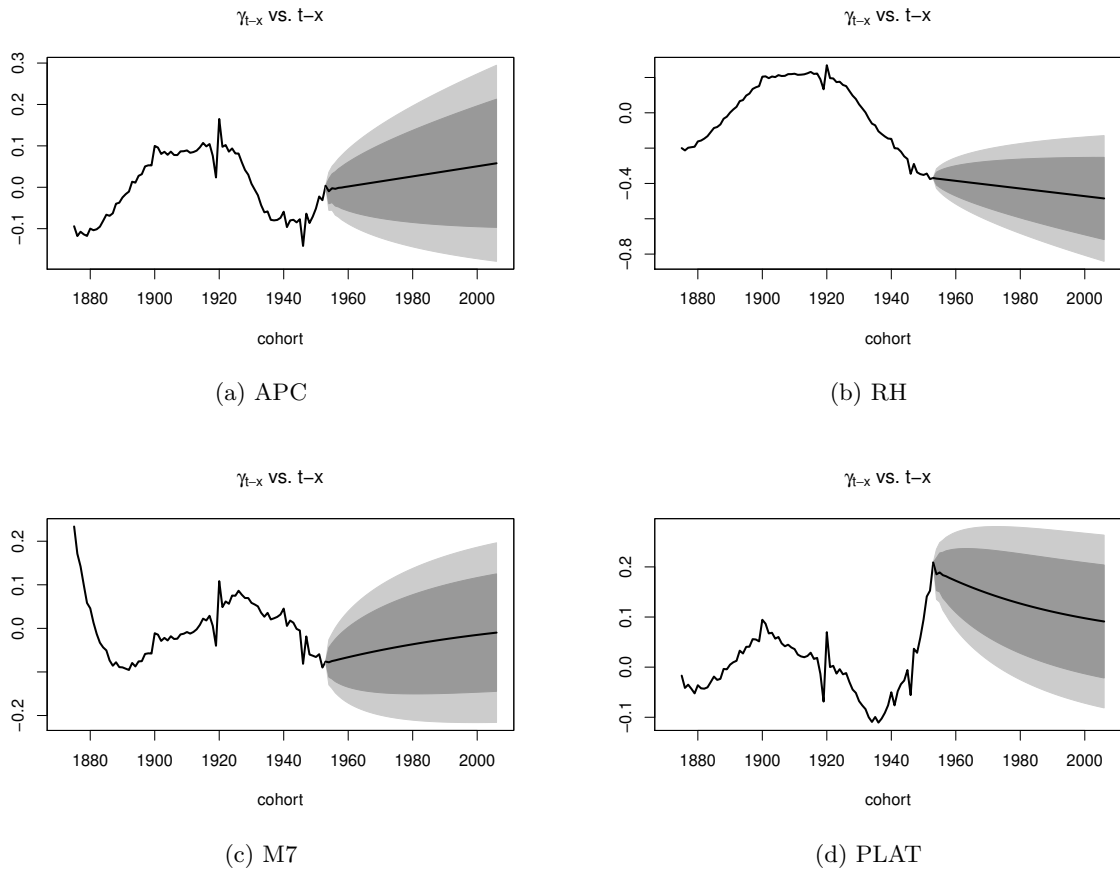[15]And the period index when Equation 20 is assumed.

Figure 8: Forecast of the cohort indexes of the APC, RH, M7 and PLAT models applied to the England and Wales males population for ages 55-89 and the period 1961-2011. Dashed lines represent central forecast and dotted lines represent 95% prediction intervals.

This output can be used to extract sample trajectories from a model. For instance, Figure 9, which depicts 20 trajectories of the period index, cohort index and one-year death probabilities at age 65 from model RH, was produced with the code:

```
R> par(mfrow = c(1, 3))
R> plot(RHfit$years, RHfit$kt[1, ], xlim = range(RHfit$years,
+    RHsim$kt.s$years), ylim = range(RHfit$kt, RHsim$kt.s$sim[1, , 1:20]),
+    type = "l", xlab = "year", ylab = "kt", main = "Period index")
R> matlines(RHsim$kt.s$years, RHsim$kt.s$sim[1, , 1:20], type = "l", lty = 1)
R> plot(RHfit$cohorts, RHfit$gc, xlim = range(RHfit$cohorts,
+    RHsim$gc.s$cohorts), ylim = range(RHfit$gc, RHsim$gc.s$sim[, 1:20],
+    na.rm = TRUE), type = "l", xlab = "year", ylab = "kt",
+    main = "Cohort index (ARIMA(1,1,0) with drift)")
R> matlines(RHsim$gc.s$cohorts, RHsim$gc.s$sim[, 1:20], type = "l", lty = 1)
R> qxt <- Dxt / Ext
R> plot(RHfit$years, qxt["65", ], xlim = range(RHfit$years, RHsim$years),
+    ylim = range(qxt["65", ], RHsim$rates["65", , 1:20]), type = "l",
```
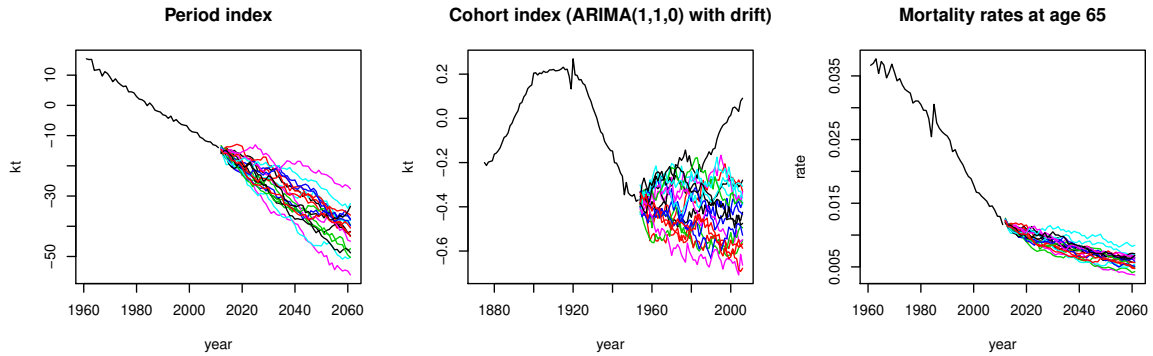
Figure 9: 20 simulated trajectories of the period index $\kappa_t^{(1)}$, cohort index γ_{t-x} and one-year death probabilities at age 65 q_{xt} for model RH fitted to the England and Wales males population for ages 55-89 and the period 1961-2011.

```
+    xlab = "year", ylab = "rate", main = "Mortality rates at age 65")
R> matlines(RHsim$years, RHsim$rates["65", , 1:20], type = "l", lty = 1)
```

We can also use the output from function `simulate` to produce fan charts depicting the uncertainty associated with a model forecast. According to Cairns *et al.* (2011), such plots are central to the analysis of the plausibility of the forecast from a model, and can be used as a criterion when deciding upon what is the most appropriate model among a group of possible stochastic mortality models. Figure 10 shows fan charts depicting 50%, 80% and 95% prediction intervals for mortality rates at ages 65, 75 and 85 for each of the six models fitted to the England and Wales experience. Such fan charts can be produced using the package **fanplot** (Abel 2015b,a). For instance, the fan chart for the Lee-Carter model shown in Figure 10a was produced with the code:

```
R> library("fanplot")
R> probs = c(2.5, 10, 25, 50, 75, 90, 97.5)
R> qxt <- Dxt / Ext
R> matplot(LCfit$years, t(qxt[c("65", "75", "85"), ]),
+    xlim = c(1960, 2061), ylim = c(0.0025, 0.2), pch = 20, col = "black",
+    log = "y", xlab = "year", ylab = "mortality rate (log scale)")
R> fan(t(LCsim$rates["65", , ]), start = 2012, probs = probs, n.fan = 4,
+    fan.col = colorRampPalette(c("black", "white")), ln = NULL)
R> fan(t(LCsim$rates["75", , ]), start = 2012, probs = probs, n.fan = 4,
+    fan.col = colorRampPalette(c("red", "white")), ln = NULL)
R> fan(t(LCsim$rates["85", , ]), start = 2012, probs = probs, n.fan = 4,
+    fan.col = colorRampPalette(c("blue", "white")), ln = NULL)
R> text(1965, qxt[c("65", "75", "85"), "1990"],
+    labels = c("x = 65", "x = 75", "x = 85"))
```

From Figure 10 we note the following:

- Whilst for models CBD, RH, M7 and PLAT the fans at age 85 are wider than at age 65 in accordance with historical evidence (see Cairns *et al.* (2011, Appendix B)), for
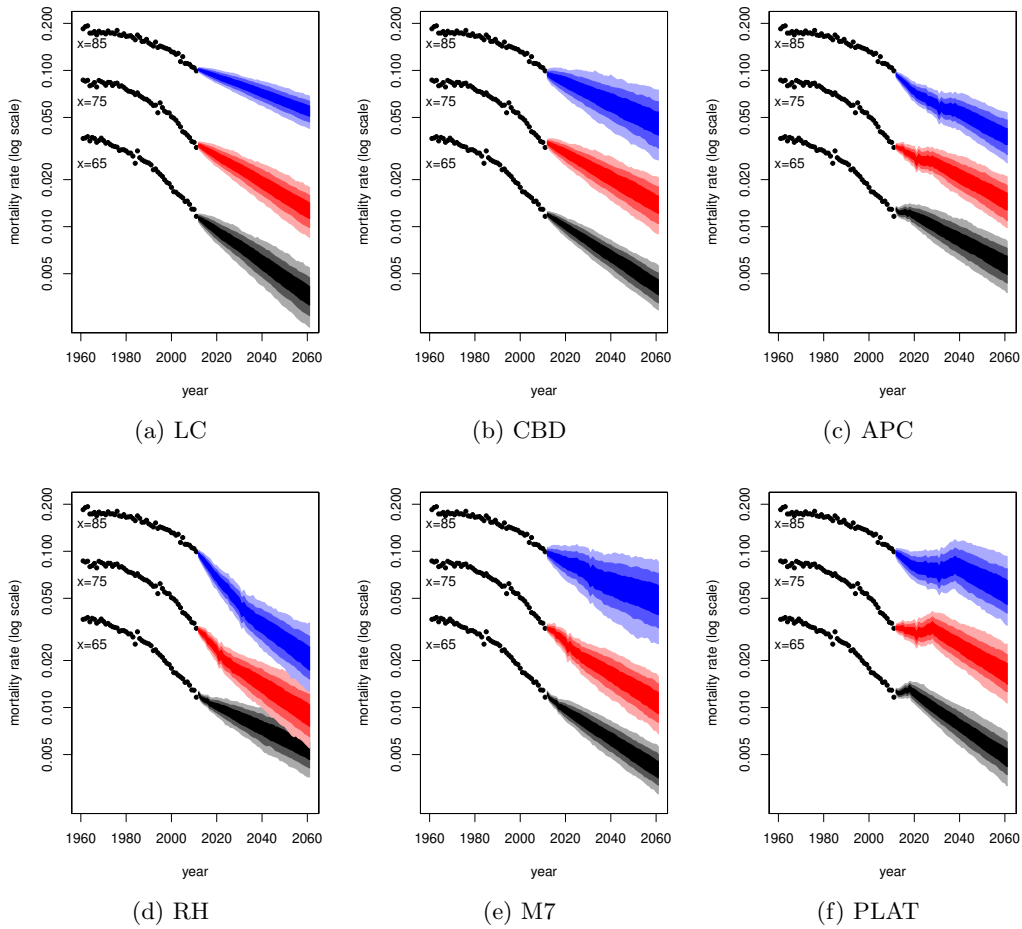
Figure 10: Fan charts for mortality rates q_{xt} at ages $x = 65$ (bottom fan), $x = 75$ (middle fan) and $x = 85$ (top fan) from the six models fitted to the England and Wales males population for ages 55-89 and the period 1961-2011. The dots show historical mortality rates for 1961-2011. Shades in the fan represent prediction intervals at the 50%, 80% and 95% level.

models LC and APC the fans at age 85 are narrower than at age 65. This suggests that forecasts from models LC and APC are not plausible for the dataset used in this paper.

- Forecasts for the PLAT model show an implausible increase of mortality rates. This is because the central trend is linked to the estimated cohort effect γ_{t-x} for the PLAT model (see Figure 8d) which shows a steep upward trend between 1935 and 1955.

- The central trend and uncertainty levels produced by the each of the models have noticeably differences. This highlights the importance of recognizing model risk as a significant issue when modeling mortality (Cairns *et al.* 2011).

Finally, with the aid of **StMoMo**'s utility function `extractCohort`, one can use the output of `forecast` and `simulate` to extract the projected death probabilities for specific cohorts. For example, a plot of the mortality rates projected by the Lee-Carter method for the 1950 cohort (see Figure 11) can be produced as follow:

Mortality rates for the 1950 cohort

Figure 11: Mortality rates for the 1950 generation obtained with the Lee-Carter model. The solid line corresponds to fitted values while the dash line corresponds to projected values.

```
R> plot(55:61, extractCohort(fitted(LCfit, type = "rates"), cohort = 1950),
+    type = "l", log = "y", xlab = "age", ylab = "q(x)",
+    main = "Mortality rates for the 1950 cohort",
+    xlim = c(55,89), ylim = c(0.005, 0.12))
R> lines(62:89, extractCohort(LCfor$rates, cohort = 1950), lty = 2)
```

These cohort mortality rates can then be employed to build a life table to perform demographic or actuarial calculations using, for instance, the **lifecontingencies** package (Spedicato 2013).

8. Parameter uncertainty and bootstrapping

When analyzing the uncertainty in mortality projections in an actuarial context it is important to consider all sources of risk. However, the prediction intervals (fan charts) obtained in the previous section only account for the uncertainty arising from the error in the forecast of the period and cohort indexes and ignore the uncertainty arising from the estimation of the parameters of the GAPC model.

Due to the analytical intractability of many stochastic mortality models, parameter uncertainty is typically accounted for using the bootstrap procedure. This procedure yields B samples $\alpha_x^b, \beta_x^{(1),b}, \ldots, \beta_x^{(N),b}, \kappa_t^{(1),b}, \ldots, \kappa_t^{(N),b}, \beta_x^{(0),b}, \gamma_{t-x}^b$, $b = 1, \ldots, B$, of the parameters of the GAPC model which can then be used to produce confidence and prediction intervals of demographic and actuarial quantities. To sample the parameters of GAPC models we consider the semiparametric bootstrap proposed by Brouhns *et al.* (2005) and the residual bootstrap first considered in Koissi *et al.* (2006):

- **Semiparametric bootstrap.** Brouhns *et al.* (2005) propose a semiparametric boot-

strap where first B samples of the number of deaths d_{xt}^b, $b = 1, \ldots, B$, are generated by sampling from the Poisson Distribution with mean d_{xt}. Each bootstrapped sample d_{xt}^b, $b = 1, \ldots, B$, is then used to re-estimate the model to obtain B bootstrapped parameter estimates $\alpha_x^b, \beta_x^{(1),b}, \ldots, \beta_x^{(N),b}, \kappa_t^{(1),b}, \ldots, \kappa_t^{(N),b}, \beta_x^{(0),b}, \gamma_{t-x}^b$, $b = 1, \ldots, B$. Renshaw and Haberman (2008) use the fitted number of death $\hat{d}_{xt}$ (instead of the observed deaths d_{xt}) to perform the sampling from the Poisson distribution. Wang and Lu (2005) consider a similar semiparametric approach using a Binomial distribution of deaths.

- **Residual bootstrap.** Another possibility is to bootstrap the residuals of the model as suggested by Koissi *et al.* (2006). Under this approach the deviance residuals r_{xt} are resampled with replacement to generate B replications r_{xt}^b, $b = 1, \ldots, B$, which are mapped to the corresponding resampled death counts d_{xt}^b, $b = 1, \ldots, B$, using the inverse formula. The model is then re-fitted using these resampled number of deaths to produce B sets of estimated parameters $\alpha_x^b, \beta_x^{(1),b}, \ldots, \beta_x^{(N),b}, \kappa_t^{(1),b}, \ldots, \kappa_t^{(N),b}, \beta_x^{(0),b}, \gamma_{t-x}^b$, $b = 1, \ldots, B$. We refer to Koissi *et al.* (2006) and Renshaw and Haberman (2008) for details on the inverse formula under a Poisson distribution of deaths and to Debón *et al.* (2010) for the inverse formula under a Binomial framework. Finally, we note that in implementing the residual bootstrap we follow Renshaw and Haberman (2008) and apply the inverse formula to produce samples of the observed number of deaths rather than samples of the fitted number of deaths as originally done by Koissi *et al.* (2006).

In what follows we illustrate the assessment of the parameter uncertainty using the package **StMoMo**. In doing so we deviate from the England and Wales example we have used so far and use instead New Zealand mortality data for males. This new example follows closely the work of Li (2014) who uses New Zealand mortality data to compare several simulation strategies for assessing the risk in mortality projections with a Poisson Lee-Carter model. The main rationale for the change of dataset is that parameter uncertainty is particularly important when analyzing the mortality of smaller populations such as smaller countries or pension plans.[16] This new example also serves as a means for illustrating the use of the **StMoMo** package with other datasets.

Mortality data for New Zealand can be extracted from the Human Mortality Database (2014) using function `hmd.mx` of the **demography** package with the code:

```
R> library("demography")
R> NZdata <- hmd.mx(country = "NZL_NP", username = username,
+    password = password, label = "New Zealand")
```

We note that the `username` and `password` above are for the Human Mortality Database and should be replaced appropriately. We can then transform the Human Mortality Database data for Kiwi males into **StMoMo**'s format using function `StMoMoData`:

```
R> NZStMoMo <- StMoMoData(NZdata, series = "male")
```

---

[16]While the population of England and Wales in 2008 was 54.8 million, the population of New Zealand in 2008 was 4.3 million.

Following Li (2014), we fit a Poisson Lee-Carter model to New Zealand male data for ages 0 to 89 and for the period 1985 to 2008. This can be carried out with the commands:[17]

```
R> LCfit_NZ <- fit(lc(), data = NZStMoMo, ages.fit = 0:89,
+    years.fit = 1985:2008)
```

In the **StMoMo** package the bootstrap of GAPC stochastic mortality models is implemented with the generic function `bootstrap`. This functions supports both the semiparametric bootstrap and the residual bootstrap. For instance, 5000 semiparametric bootstrap samples of the Lee-Carter model can be obtained with the code:

```
R> LCboot_NZ <- bootstrap(LCfit_NZ, nBoot = 5000, type = "semiparametric")
```

We note that the bootstrap is a computationally intensive procedure. In particular, the 5000 semiparametric bootstrap samples of the Lee-Carter model took about two hours to run.[18]

The output from function `bootstrap` is an object of the class `"bootStMoMo"` in which the component `bootParameters` contains the `nBoot` replications of the bootstrap parameters. A fan chart depicting the 50%, 80% and 95% confidence intervals of the bootstrapped Lee-Carter model (Figure 12) can be obtained with the command:

```
R> plot(LCboot_NZ, nCol = 3)
```

From Figure 12 we note that whilst the parameter uncertainty in the static age function α_x and the period index $\kappa_t^{(1)}$ is modest, the uncertainty in the age-modulating parameters $\beta_x^{(1)}$ is more significant.

Once a stochastic mortality model has been bootstrapped we can simulate it forward to obtain simulated trajectories which account for both the forecast error in the period and cohort indexes and the error in the model fitting. In **StMoMo** we can accomplish this using the function `simulate` applied to an object of class `"bootStMoMo"`. Thus, to obtain 5000 simulated trajectories of the Lee-Carter model for the next 24 years taking into account parameter uncertainty we use the instruction:

```
R> LCsimPU_NZ <- simulate(LCboot_NZ,  h = 24)
```

To highlight the impact of parameter risk on mortality rate projections it is instructive to compare prediction intervals with and without the allowance of parameter uncertainty. 24-year ahead central forecast together with 5000 trajectories of the Lee-Carter model allowing only for forecast error in the random walk with drift and ignoring the model fitting error can be obtained with the code:

---

[17]We note that fitting a Poisson Lee-Carter model can also be done using function `lca` in **demography** by setting argument `adjust="dxt"`. Similarly, we can a fit a Poisson Lee-Carter model using function `lca.rh` in package **ilc**.

[18]In running this code, we used a computer with an Intel Core i5-3320m processor running at 2.60 GHz under Windows 7 Home Premium Edition (64 bits) with 8 GB of RAM.
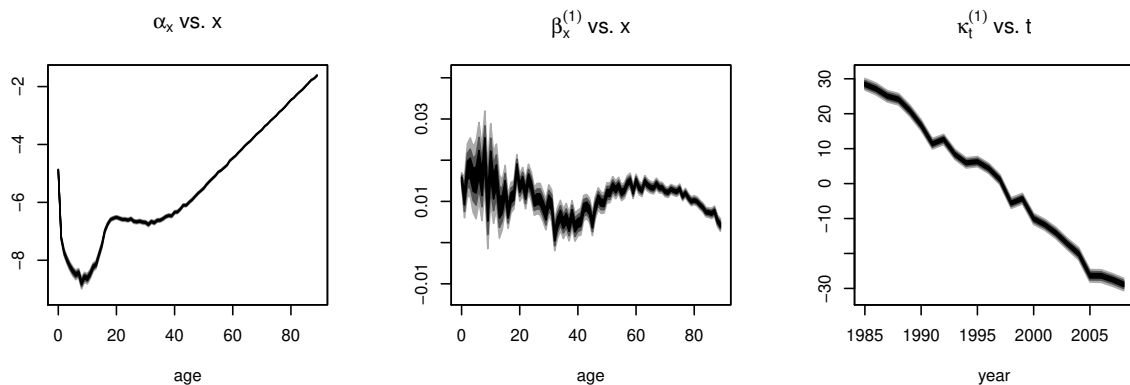
Figure 12: Bootstrapped parameters for the Poisson Lee-Carter model fitted to the New Zealand male population for ages 0-89 and the period 1985-2008. Shades in the fan represent confidence intervals at the 50%, 80% and 95% level.

```
R> LCfor_NZ <- forecast(LCfit_NZ, h = 24)
R> LCsim_NZ <- simulate(LCfit_NZ, nsim = 5000, h = 24)
```

Figure 13 depicts 95% prediction intervals for mortality rates at age 40, 60 and 80 with and without allowance for parameter uncertainty. This graph was produced with the code:

```
R> mxt <- LCfit_NZ$Dxt / LCfit_NZ$Ext
R> mxtHat <- fitted(LCfit_NZ, type = "rates")
R> mxtCentral <- LCfor_NZ$rates
R> mxtPred2.5 <- apply(LCsim_NZ$rates, c(1, 2), quantile, probs = 0.025)
R> mxtPred97.5 <- apply(LCsim_NZ$rates, c(1, 2), quantile, probs = 0.975)
R> mxtHatPU2.5 <- apply(LCsimPU_NZ$fitted, c(1, 2), quantile, probs = 0.025)
R> mxtHatPU97.5 <- apply(LCsimPU_NZ$fitted, c(1, 2), quantile, probs = 0.975)
R> mxtPredPU2.5 <- apply(LCsimPU_NZ$rates, c(1, 2), quantile, probs = 0.025)
R> mxtPredPU97.5 <- apply(LCsimPU_NZ$rates, c(1, 2), quantile, probs = 0.975)
R> x <- c("40", "60", "80")
R> matplot(LCfit_NZ$years, t(mxt[x, ]), xlim = range(LCfit_NZ$years,
+    LCfor_NZ$years), ylim = range(mxtHatPU97.5[x, ], mxtPredPU2.5[x, ],
+    mxt[x, ]), type = "p", xlab = "years",
+    ylab = "mortality rates (log scale)", log = "y", pch = 20,
+    col = "black")
R> matlines(LCfit_NZ$years, t(mxtHat[x, ]), lty = 1, col = "black")
R> matlines(LCfit_NZ$years, t(mxtHatPU2.5[x, ]), lty = 5, col = "red")
R> matlines(LCfit_NZ$years, t(mxtHatPU97.5[x, ]), lty = 5, col = "red")
R> matlines(LCfor_NZ$years, t(mxtCentral[x, ]), lty = 4, col = "black")
R> matlines(LCsim_NZ$years, t(mxtPred2.5[x, ]), lty = 3, col = "black")
R> matlines(LCsim_NZ$years, t(mxtPred97.5[x, ]), lty = 3, col = "black")
R> matlines(LCsimPU_NZ$years, t(mxtPredPU2.5[x, ]), lty = 5, col = "red")
R> matlines(LCsimPU_NZ$years, t(mxtPredPU97.5[x, ]), lty = 5, col = "red")
R> text(1986, mxtHatPU2.5[x, "1995"], labels = c("x=40", "x=60", "x=80"))
```
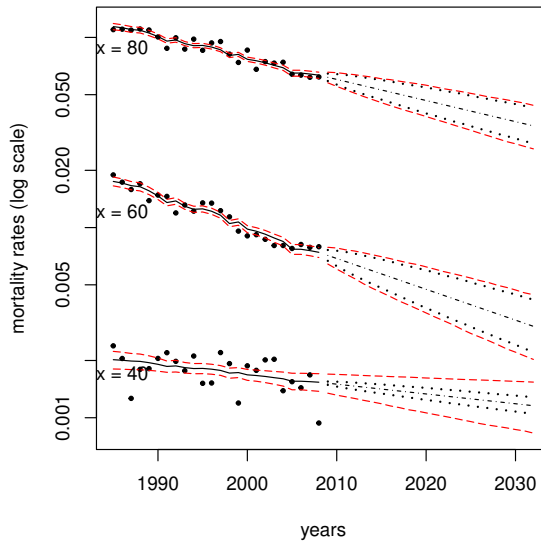
Figure 13: 95% Prediction intervals for mortality rates q_{xt} at ages $x = 40$ (bottom lines), $x = 60$ (middle lines) and $x = 80$ (top lines) for the Poisson Lee-Carter model fitted to the New Zealand male population for ages 0-89 and the period 1985-2008. Dots show historical mortality rates for 1985-2008 and solid black lines show the corresponding fitted rates. Dashed lines represent central forecast, black dotted lines represent 95% prediction intervals excluding parameter uncertainty and dot-dashed red lines depict 95% confidence and prediction intervals including parameter uncertainty.

In Figure 13 we can clearly see that parameter uncertainty has an important impact on the prediction intervals. This is particularly noticeable at age 40 where in year 2030, for instance, the width of the prediction interval with parameter uncertainty is around 3 times bigger than that without parameter uncertainty. These results are in line with those obtained by Li (2014) using the same dataset.[19]

9. Conclusion

In this paper we have introduced the family of generalized age-period-cohort stochastic mortality models by paralleling the standard framework of generalized linear models. In addition, we have presented the R package **StMoMo** which takes advantage of the unifying framework of the GAPC family to provide tools for fitting a diverse number of stochastic mortality models, assessing their goodness of fit and also performing mortality projections. A key feature of the GAPC family and of **StMoMo** is that they not only encompass models from the Lee-Carter and CBD families, but can also accommodate possible new models. Furthermore, model risk is a prevalent issue when forecasting mortality and we therefore believe that the possibility of easily implementing and comparing a wide range of models makes **StMoMo** a valuable addition to the toolkit for measuring and managing longevity risk.

---

[19] We note that our prediction intervals without and with parameter uncertainty correspond to methods (I) and (IVa) in Li (2014), respectively.

Our package can be expanded in several directions. The current version of **StMoMo** only allows the use of a log link with a Poisson distribution of deaths or a logit link with a Binomial distribution of deaths. However, we plan to expand the possible combinations of error distribution and link function to include, for instance, Binomial errors with a complementary log-log link as suggested by Currie (2016).

In the GAPC family it is assumed that the age-modulating terms $\beta_x^{(i)}$, $i = 0, \ldots, N$, are either non-parametric functions of age which need to be estimated or parametric functions of age with a pre-specified functional form $f^{(i)}(x)$. This latter case could be extended to include the more general case of a pre-specified functional form with a set of parameters that need to be estimated, that is, $\beta_x^{(i)} = f^{(i)}(x; \theta_i)$ with θ_i being some model parameters. This generalization would allow the implementation of the family of models considered in Hunt and Blake (2014) and the inclusion of models where age terms are given by smooth parametric functions such as polynomial splines. Smoothness of the parameters in a stochastic mortality model is a topic that has received attention in the literature and offers a potential route for further extensions of **StMoMo**, see for instance Delwarde *et al.* (2007); Currie (2013).

Finally, the increasing attention that multiple population mortality models are receiving in the mortality forecasting literature provides an avenue for a wealth of extensions of our package. In particular, we are considering the development of a two-population version of **StMoMo** that implements the class of relative two-population models considered in Villegas *et al.* (2015).

Acknowledgements

References

Abel GJ (2015a). "fanplot: An R Package for Visualising Sequential Distributions." *The R Journal*, **7**(2), 15–23. URL http://journal.r-project.org/archive/2015-1/abel.pdf.

Abel GJ (2015b). ***fanplot****: Visualisation of Sequential Probability Distributions Using Fan Charts*. R package version 3.4.0, URL http://CRAN.R-project.org/package=fanplot.

Alai DH, Sherris M (2014). "Rethinking Age-Period-Cohort Mortality Trend Models." *Scandinavian Actuarial Journal*, (3), 208–227.

Aro H, Pennanen T (2011). "A User-Friendly Approach to Stochastic Mortality Modelling." *European Actuarial Journal*, **1**, 151–167.

Booth H, Hyndman RJ, Tickle L (2014). "Prospective Life Tables." In A Charpentier (ed.), *Computational Actuarial Science with R*, pp. 319–344. CRC Press, Boca Raton, FL.

Booth H, Maindonald J, Smith L (2002). "Applying Lee-Carter under Conditions of Variable Mortality Decline." *Population Studies*, **56**(3), 325–336.

Börger M, Fleischer D, Kuksin N (2013). "Modeling the Mortality Trend under Modern Solvency Regimes." *ASTIN Bulletin*, **44**(1), 1–38.

Brouhns N, Denuit M, Van Keilegom I (2005). "Bootstrapping the Poisson Log-Bilinear Model for Mortality Forecasting." *Scandinavian Actuarial Journal*, (3), 212–224.

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

Butt Z, Haberman S, Shang HL (2014). ***ilc****: Lee-Carter Mortality Models Using Iterative Fitting Algorithms*. R package version 1.0, URL http://CRAN.R-project.org/package=ilc.

Cairns AJG, Blake D, Dowd K (2006). "A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration." *Journal of Risk and Insurance*, **73**(4), 687–718.

Cairns AJG, Blake D, Dowd K, Coughlan GD, Epstein D, Khalaf-Allah M (2011). "Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models." *Insurance: Mathematics and Economics*, **48**(3), 355–367.

Cairns AJG, Blake D, Dowd K, Coughlan GD, Epstein D, Ong A, Balevich I (2009). "A Quantitative Comparison of Stochastic Mortality Models Using Data from England and Wales and the United States." *North American Actuarial Journal*, **13**(1), 1–35.

Camarda CG (2012). "**MortalitySmooth**: An R Package for Smoothing Poisson Counts with P-Splines." *Journal of Statistical Software*, **50**(1), 1–24.

Clayton D, Schifflers E (1987). "Models for Temporal Variation in Cancer Rates. II: Age-Period-Cohort Models." *Statistics in Medicine*, **6**(4), 469–481.

Currie ID (2006). "Smoothing and Forecasting Mortality Rates with P-Splines." URL http://www.macs.hw.ac.uk/{~}iain/research/talks/Mortality.pdf.

Currie ID (2013). "Smoothing constrained generalized linear models with an application to the Lee-Carter model." *Statistical Modelling*, **13**(1), 69–93.

Currie ID (2016). "On Fitting Generalized Linear and non-linear Models of Mortality." *Scandinavian Actuarial Journal*, (4), 356–383.

Currie ID, Durban M, Eilers PHC (2006). "Generalized linear array models with applications to multidimensional smoothing." *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, **68**(2), 259–280.

Debón A, Martínez-Ruiz F, Montes F (2010). "A Geostatistical Approach for Dynamic Life Tables: The Effect of Mortality on Remaining Lifetime and Annuities." *Insurance: Mathematics and Economics*, **47**(3), 327–336.

Delwarde A, Denuit M, Eilers P (2007). "Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach." *Statistical Modelling*, **7**(1), 29–48.

Haberman S, Kaishev VK, Millossovich P, Villegas AM, Baxter S, Gaches A, Gunnlaugsson S, Sison M (2014). "Longevity Basis Risk: A Methodology for Assessing Basis Risk." *Institute and Faculty of Actuaries Sessional Research Paper*. URL http://www.actuaries.org.uk/sites/all/files/documents/pdf/ifoa-llma-longevity-basis-risk-report.pdf.

Haberman S, Renshaw A (2009). "On Age-Period-Cohort Parametric Mortality Rate Projections." *Insurance: Mathematics and Economics*, **45**(2), 255–270.

Haberman S, Renshaw A (2011). "A Comparative Study of Parametric Mortality Projection Models." *Insurance: Mathematics and Economics*, **48**(1), 35–55.

Hobcraft J, Menken J, Preston S (1982). "Age, Period, and Cohort Effects in Demography: A Review." *Population Index*, **48**(1), 4–43.

Human Mortality Database (2014). "University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)." URL www.mortality.org.

Hunt A, Blake D (2014). "A General Procedure for Constructing Mortality Models." *North American Actuarial Journal*, **18**(1), 116–138.

Hunt A, Blake D (2015). "On the Structure and Classification of Mortality Models Mortality Models." *Pension Institute Working Paper*. URL http://www.pensions-institute.org/workingpapers/wp1506.pdf.

Hunt A, Villegas AM (2015). "Robustness and Convergence in the Lee-Carter Model with Cohorts." *Insurance: Mathematics and Economics*, **64**, 186–202.

Hyndman RJ (2017). **forecast**: *Forecasting Functions for Time Series and Linear Models*. R package version 8.0, URL http://github.com/robjhyndman/forecast.

Hyndman RJ, Booth H, Tickle L, Maindonald J (2014). **demography**: *Forecasting Mortality, Fertility, Migration and Population Data*. R package version 1.18, URL http://CRAN.R-project.org/package=demography.

Hyndman RJ, Khandakar Y (2008). "Automatic Time Series Forecasting: The **forecast** Package for R." *Journal of Statistical Software*, **27**(3), 1–22. URL http://www.jstatsoft.org/v27/i03/.

Hyndman RJ, Ullah S (2007). "Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach." *Computational Statistics & Data Analysis*, **51**(10), 4942–4956.

Koissi M, Shapiro A, Hognas G (2006). "Evaluating and Extending the Lee-Carter Model for Mortality Forecasting: Bootstrap Confidence Interval." *Insurance: Mathematics and Economics*, **38**(1), 1–20.

Lee R, Miller T (2001). "Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality." *Demography*, **38**(4), 537–549.

Lee RD, Carter LR (1992). "Modeling and Forecasting U.S. Mortality." *Journal of the American Statistical Association*, **87**(419), 659–671.

Li J (2014). "A Quantitative Comparison of Simulation Strategies for Mortality Projection." *Annals of Actuarial Science*, **8**(2), 281–297.

Lovász E (2011). "Analysis of Finnish and Swedish Mortality Data with Stochastic Mortality Models." *European Actuarial Journal*, **1**(2), 259–289.

McCullagh P, Nelder J (1989). *Generalized Linear Models.* second edition. Chapman & Hal, London.

Oeppen J, Vaupel JW (2002). "Broken Limits to Life Expectancy." *Science*, **296**(5570), 1029–1031.

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

Osmond C (1985). "Using age, period and cohort models to estimate future mortality rates." *International Journal of Epidemiology*, **14**(1), 124–129. ISSN 03005771. `doi:10.1093/ije/14.1.124`.

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

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Renshaw A, Haberman S (2003). "Lee-Carter Mortality Forecasting with Age-specific Enhancement." *Insurance: Mathematics and Economics*, **33**(2), 255–272.

Renshaw A, Haberman S (2006). "A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors." *Insurance: Mathematics and Economics*, **38**(3), 556–570.

Renshaw A, Haberman S (2008). "On Simulation-Based Approaches to Risk Measurement in Mortality with Specific Reference to Poisson Lee-Carter Modelling." *Insurance: Mathematics and Economics*, **42**(2), 797–816.

Renshaw A, Haberman S, Hatzopoulos P (1996). "The Modelling of Recent Mortality Trends in United Kingdom Male Assured Lives." *British Actuarial Journal*, **2**(2), 449–477.

Spedicato GA (2013). "The lifecontingencies Package : Performing Financial and Actuarial Mathematics Calculations in R." *Journal of Statistical Software*, **55**(10).

Turner H, Firth D (2015). *Generalized Nonlinear Models in R: An Overview of the* **gnm** *Package.* R package version 1.0-8, URL http://CRAN.R-project.org/package=gnm.

van Berkum F, Antonio K, Vellekoop M (2014). "The Impact of Multiple Structural Changes on Mortality Predictions." *Scandinavian Actuarial Journal.*

Villegas AM, Haberman S, Kaishev V, Millossovich P (2015). "A Comparative Study of Two-Population Models for the Assessment of Basis Risk in Longevity Hedges." *Working paper.*

Villegas AM, Millossovich P, Kaishev V (2017). **StMoMo**: *An R Package for Stochastic Mortality Modelling.* R package version 0.4.0, URL http://CRAN.R-project.org/package=StMoMo.

Wang D, Lu P (2005). "Modelling and Forecasting Mortality Distributions in England and Wales Using the Lee–Carter Model." *Journal of Applied Statistics*, **32**(9), 873–885.

Wikipedia (2014). "King Momo, in Wikipedia, The Free Encyclopedia." URL http://en.wikipedia.org/w/index.php?title=King{_}Momo{&}oldid=635124783.

Willets R (2004). "The Cohort Effect: Insights and Explanations." *British Actuarial Journal*, **10**(4), 833–877.

Affiliation:

Andrés M. Villegas
School of Risk and Actuarial Studies and
ARC Centre of Excellence in Population Ageing Research (CEPAR)
UNSW Business School
UNSW Sydney 2052, Australia
E-mail: a.villegas@unsw.edu.au

Pietro Millossovich
Faculty of Actuarial Science and Insurance
Cass Business School
City, University of London
106 Bunhill Row, London, EC1Y 8TZ, UK
E-mail: Pietro.Millossovich.1@city.ac.uk

Vladimir K. Kaishev
Faculty of Actuarial Science and Insurance
Cass Business School
City, University of London
106 Bunhill Row, London, EC1Y 8TZ, UK
E-mail: Vladimir.Kaishev.1@city.ac.uk