On the Structure and Classification of Mortality Models

Andrew Hunt*
Cass Business School, City University London
Corresponding author: andrew.hunt.1@cass.city.ac.uk†

David Blake
Pensions Institute, Cass Business School, City University London

December 2017

Abstract

There has recently been a huge increase in the use of models which examine the structure of mortality rates across the dimensions of age, period and cohort. This paper reviews the major developments in the field and provides a holistic analysis of these models and examines their similarities and differences. Specifically, it reviews the models that have been proposed to date, investigates the structure of age/period/cohort mortality models, introduces a classification scheme for existing models and lists the key principles a model user should consider when constructing a new model in this class.
JEL Classification: C33, C35, C51
Keywords: Mortality modelling, age/period/cohort models, age period effects, cohort effects

*We are grateful to Andrés Villegas, Steven Haberman, Bent Nielsen, Andrew Cairns, Pietro Millossovich and Ana Debón for many useful discussions on this and related topics which have improved this paper immeasurably.

†All comments and criticisms are gratefully received and we hope to be able to respond to them in further revisions of this paper.

1 Introduction

Recent years have witnessed a dramatic increase in the attention paid to the study of the evolution and projection of mortality rates. Demographers, statisticians and actuaries across the world have woken up to the issues caused by rising longevity and aging populations.

Much of the analysis of the historical evolution of mortality rates is made using models which decompose mortality rates across the dimensions of age, period and cohort (or year of birth). These three variables form a natural way of analysing how mortality rates change for individuals as they age, the impact of medical and social progress with time, and the lifelong mortality effects which follow individuals from birth. By projecting the effects of period and cohort, we can also gain insights into the likely path mortality rates might take in future.

Since the number of age/period/cohort (APC) models has increased rapidly in recent years, we believe that the time has come to first review these developments and then undertake a more holistic analysis of APC models. This paper reviews the major contributions to the development of new APC mortality models, before analysing the common structure of these models and proposing a way of classifying the models proposed to date. It also seeks to assess the key principles a model user should consider before selecting or constructing a model appropriate to their aims. While most of the issues raised in this paper will be familiar to many model users, we believe that a proper understanding of the structure of APC models is needed in order to avoid using a poorly specified model. For instance, many of the issues raised and pitfalls identified in these papers were vital to the development of the "general procedure" for constructing APC mortality models, described in Hunt and Blake (2014).

The format of this paper is to discuss the basic structure of the majority of APC models which have been proposed to date in Section 2. We then review the mortality models that adopt this structure in Section 3. The components of this structure are further discussed in terms of

- the connections between the data, the variables of interest and our predictor structure in Section 4;

- the inclusion of a static function of age in Section 5;
- the potential forms for the dynamic structure across ages in the model in Section 6; and
- the issues raised by the inclusion of parameters to capture the effects of year of birth in the data and how these can be resolved in Section 7.

Section 8 provides a simple classification of APC models that highlights the key decisions which have to be made in order to select the most suitable model for the task at hand. Finally, we draw conclusions in Section 9.

2 Age/period/cohort structure

An APC mortality model is one which links a response variable with a linear or bilinear predictor structure consisting of a series of factors dependent on age, x, period, t, and year of birth (or cohort), $y = t - x$, for a population. APC models therefore fit into the general class of generalised non-linear models (GNMs), with a structure that can be written as follows:

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

Most models of this form can be fitted to data using the **StMoMo** package in R (Villegas et al. (2015)), which takes advantage of the common GNM structure to efficient estimate a wide range of different models.

This structure has the following components:

- A link function, $\eta_{x,t}$, to transform the response variable (which will be some measure of mortality rates) at age x and for year t into a form suitable for modelling and link it to the proposed predictor structure.
- A static age function, α_x, to capture the general shape of mortality across all ages and features of the mortality curve which do not change with time.
- A set of N age/period terms, $\beta_x^{(i)} \kappa_t^{(i)}$, consisting of period functions, $\kappa_t^{(i)}$, determining the evolution of mortality rates through time, and

age functions, $\beta_x^{(i)}$, determining the pattern of mortality change across ages. The choice of suitable forms for the age functions is discussed in Section 6.

- An age/cohort term, $\beta_x^{(0)}\gamma_{t-x}$, consisting of a cohort term, γ_{t-x}, which determines the lifelong effects specific to each generation, denoted by their year of birth, and an age function, $\beta_x^{(0)}$, which modifies the cohort term.[1]

Each of these component terms is discussed in greater detail in the sections below. One advantage of most APC mortality models is that the components in them can be interpreted in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates which generate them. We call such an interpretation the "demographic significance" of each term. Demographic significance is, by definition, subjective, since it relates to the interpretation of the parameters. However, it is still a useful concept as it motivates many of the decisions around the construction of mortality models and their projection into the future.

While this structure is not exhaustive, it does encompass the vast majority of the discrete-time mortality models which have been proposed to date. In particular, it is worth noting that we have assumed that the period functions can vary freely for each year and are not constrained to be smooth functions. This is the key feature which enables these models to be projected stochastically and therefore generate probabilistic forecasts of future mortality rates.

In contrast, some models require that the period functions be modelled through a series of deterministic functions and so are projected by extrapolating these functions into the future. For example, the P-splines model proposed in Currie et al. (2004) and the model of Sithole et al. (2000), which might otherwise be felt to fit into the APC framework, require that the period functions are cubic b-splines and Legendre polynomials, respectively. This typically restricts the application of these models to smoothing historical data or short-term projections of mortality. We therefore do not consider

[1]Most APC mortality models have only one age/cohort term for the reasons discussed in Section 7. However, some models do incorporate multiple terms, for instance, that proposed in Hatzopoulos and Haberman (2011).

these models further in this paper.

Recently, a number of papers such as Mitchell et al. (2013) and Haberman and Renshaw (2012, 2013) have modified the structure in Equation 1 to model mortality improvement rates rather than the mortality rates themselves. The different interpretations placed on the response variables of interest and terms within the predictor structure make mortality improvement models qualitatively different from the class of models considered within this paper, and so we do not discuss these models further.

Finally, it is worth noting that the predictor structure in Equation 1 could also be extended to include a range of explanatory variables which might influence mortality rates. These regressors might include variables relating to the health of the population (for instance, smoking prevalence was considered in Wang and Preston (2009) and Kleinow and Cairns (2013)) or macroeconomic variables such as GDP growth or unemployment (e.g., Reichmuth and Sarferaz (2008) and Hanewald (2011)). Such an approach is a natural way of modelling the underlying drivers of changing mortality and highlights the flexibility of the APC approach, but, again, is not considered further in this paper.

3 A brief history of APC mortality models

APC mortality models have sometimes been classified into three broad families, for instance in Cairns (2014). These can be thought of as:

1. the Lee-Carter (LC) model, proposed in Lee and Carter (1992), and its direct extensions, discussed in Section 3.1;

2. the Cairns-Blake-Dowd (CBD) model, proposed in Cairns et al. (2006a), and its direct extensions, which we discuss in Section 3.2; and

3. the classic APC model of Hobcraft et al. (1982) and others, and the hybridisation of this model with those from the other two families, as discussed in Section 3.3.

Although we do not consider this to be a particularly useful classification for the purposes of understanding the properties of the specific mortality models,

it is helpful in understanding the evolution of the field of APC mortality modelling as it has developed over recent decades.

3.1 Lee-Carter models

The first and still the most widely used mortality model is that proposed in Lee and Carter (1992). The LC model has a single age/period term of the form

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t \tag{2}$$

and models the logarithm of the force of mortality, $\mu_{x,t}$.[2] In Lee and Carter (1992), this model was fitted using a two-stage process: the first stage estimated the parameters using singular value decomposition (SVD), which is an application of least squares fitting methods for a bilinear predictor structure, and the second stage adjusted κ_t in order to better fit the observed number of deaths in each year. Other studies, for example, Wilmoth (1993) and Lee (2000), adapted the two-stage approach but retained the use of least squares estimation. However, superior fitting methods based on maximum likelihood estimation have been proposed in Brouhns et al. (2002), although these have not been universally adopted.

A great deal of the early work focused on developing the fitting methods used for the model, such as Wilmoth (1993), Lee and Miller (2001) and Renshaw and Haberman (2003a,c), and investigating the stability of the parameter estimates for different ranges of data (for instance, Carter and Prskawetz (2001) and Booth et al. (2002)). However, it was Booth et al. (2002) which first highlighted that the use of SVD to fit the LC model selects only the first of a potentially large number of age/period terms. Therefore, the LC model can easily be directly extended to more complicated models of the form

$$\ln(\mu_{x,t}) = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} \tag{3}$$

[2]Lee and Carter (1992) originally applied the LC model to central mortality rates, $m_{x,t}$. However, as discussed in Section 4, these are equivalent to the force of mortality, $\mu_{x,t}$, under the assumption in Equation 23.

These models were not directly investigated in Booth et al. (2002), but the two-term model (commonly referred to as the LC2 model)

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} \tag{4}$$

was studied in detail in Renshaw and Haberman (2003b). A variation of this model, where $\kappa_t^{(1)}$ is replace by a linear function in time, i.e.,

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x^{(1)}(t - \bar{t}) + \beta_x^{(2)}\kappa_t \tag{5}$$

where $\bar{t} = \frac{1}{T}\sum_t t$, was studied in detail in Callot et al. (2014). Furthermore, a three-term extension of the LC model was investigated for Hungarian mortality data in Baran et al. (2007).

Subsequently, models with multiple age/period terms were studied in Hyndman and Ullah (2007), which used functional data analysis to fit the models to data, Hatzopoulos and Haberman (2009), which used generalised linear models, and Wang et al. (2009), which used principal components analysis. However, these extensions to the LC model have proved less popular in practice than might have been expected, possibly because the higher order period functions show complicated behaviour, with apparent changes in trend, which make forecasting difficult.

3.2 Cairns-Blake-Dowd models

One of the most popular competitor models to the LC model was introduced in Cairns et al. (2006a) in order to overcome the problem that projected mortality rates are perfectly correlated in single age/period term models. The CBD model is usually written in the form

$$\text{logit}(q_{x,t}) = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} \tag{6}$$

where $\text{logit}(z) \equiv \ln(z) - \ln(1 - z)$ and $\bar{x}$ denotes the average age in the data range being used. In addition to the multiple age/period terms, Cairns et al. (2006a) highlighted the absence of the static age function, α_x, (which reduces the number of free parameters and so obtains a more parsimonious model) and the use of the logit-link function as key differentiators of the CBD approach from the preceding LC-style models. In Cairns et al. (2006a), the model was fitted to data by using ordinary least squares on the transformed

probabilities of death, $q_{x,t}$.

Since the age functions in the CBD model take the form of a constant and a linear function of age, respectively, it is natural to extend the model by adding age/period terms with higher-order polynomial age functions, for instance

$$\text{logit}(q_{x,t}) = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + ((x - \bar{x})^2 - \sigma_x)\kappa_t^{(3)} \tag{7}$$

where $\sigma_x = \frac{1}{X}\sum_x (x - \bar{x})^2$. However, such a "quadratic CBD" model has not been used directly until Yang et al. (2015). Instead, in Cairns et al. (2009), a general form of CBD-style models is proposed

$$\text{logit}(q_{x,t}) = \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} \gamma_{t-x}^{(i)} \tag{8}$$

This form of APC mortality models is more general than that described in Section 2, since it allows for "trilinear" terms containing interactions between age, period and cohort. However, in Cairns et al. (2009), the authors only investigate the models

$$\text{logit}(q_{x,t}) = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \gamma_{t-x} \tag{9}$$

$$\text{logit}(q_{x,t}) = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + ((x - \bar{x})^2 - \sigma_x)\kappa_t^{(3)} + \gamma_{t-x} \tag{10}$$

$$\text{logit}(q_{x,t}) = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + (x - x_c)\gamma_{t-x} \tag{11}$$

which are referred to as models M6, M7 and M8, respectively. All of these explicitly extend either the original or the quadratic CBD model with a cohort term. M8 goes further and introduces a linear age function, $x - x_c$, to the cohort effect, where x_c gives the age where cohort effects are weakest. However, in practice, M7 has proved the most popular extension of the original CBD model, since it gives a better fit to data than M6 (which it nests) and the age function for the cohort parameters in M8 may be more complicated to fit to data due to the need to select appropriate x_c.

A similar, but not explicitly related model, was proposed in Aro and Pennanen (2011). This took the form

$$\text{logit}(p_{x,t}) = \sum_{i=1}^{N} f^{(i)}(x)\kappa_t^{(i)} \tag{12}$$

where the age functions, $f^{(i)}(x)$, were piecewise linear splines chosen to be orthogonal to each other (i.e., $\sum_x f^{(i)}(x)f^{(j)}(x) = 0$, for $i \neq j$) and $p_{x,t}$ is the one-year probability of survival for age x and year t. This model was fitted to data using binomial maximum likelihood estimation. Aro and Pennanen (2011) investigated this model with either two or three age/period terms. However, since $p_{x,t} = 1 - q_{x,t}$, we see that $\text{logit}(p_{x,t}) = -\text{logit}(q_{x,t})$ and, therefore, the two-factor model proposed in Aro and Pennanen (2011) can be rearranged to be mathematically equivalent to the original CBD model in Equation 6.

3.3 The classic APC model and hybrid models

The final broad family of mortality model are those based on the classic APC model of Hobcraft et al. (1982), Osmond (1985), O'Brien (2000), Carstensen (2007) and Kuang et al. (2008b) (amongst others). In its original form, this model is written

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t + \gamma_{t-x} + \delta$$

which is symmetric in each of the three dimensions of age, period and cohort. However, when used for modelling mortality rates, this is usually simplified to

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t + \gamma_{t-x} \tag{13}$$

by combining α_x and δ in order to be more comparable with LC-style models. This model was often compared to the even simpler age/period and age/cohort models

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t$$
$$\ln(\mu_{x,t}) = \alpha_x + \gamma_{t-x}$$

in order to determine whether the effects of period or cohort were more significant in the data being analysed. The age/period model can be naturally compared to the LC model with $\beta_x = 1$, while the age/cohort model has been suggested in the context of modelling longevity risk in Alai and Sherris (2012).

Traditionally, the classic APC model was fitted using ordinary least squares, since it is linear in the parameters. However, this involves the construction

of a "design matrix", which is singular and therefore cannot be inverted as required by the least squares approach. This singularity of the design matrix, caused by the lack of identifiability of the parameters in the classic APC model, has generated a substantial literature on different methods for fitting the model, as discussed in Glenn (1976), Fienberg and Mason (1979), Rodgers (1982), Holford (1983), Clayton and Schifflers (1987), Wilmoth (1990), Yang et al. (2004), Kuang et al. (2008a) and O'Brien (2011). See also Hunt and Blake (2015), which looks at the identifiability issues in this model and more complicated APC mortality models.

The classic APC model was directly extended in Wilmoth (1990) to give

$$\ln\left(\frac{q_{x,t}}{1-0.5q_{x,t}}\right) = \alpha_x + \kappa_t^{(1)} + \sum_{i=2}^{N} \beta_x^{(i)}\kappa_t^{(i)} + \gamma_{t-x} \tag{14}$$

where the higher order age/period terms were found using a weighted least squares approach. Although this model gave a relatively good fit to data, it has not been widely used for projecting mortality rates.

Despite the classic APC model having a very long history in the fields of medicine, epidemiology and sociology, it was little used for the purposes of measuring longevity risk until the work of Willets (1999, 2004) brought a greater focus onto potential cohort effects in population data. The classic APC model was subsequently introduced into actuarial science in Currie (2006), which also proposed smoothing the terms in it using P-splines to reduce the number of free parameters.

The introduction of the classic APC model for use in projecting mortality rates has generated a significant number of models which seek to combine its features with those of models from the LC and CBD families, in order to create hybrid models. Most obviously, the cohort term was added directly to the CBD family in models M6, M7 and M8 discussed in Section 3.2. Cohort terms were also added to the LC family of models in Renshaw and Haberman (2006) and Haberman and Renshaw (2009) to give

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x^{(1)}\kappa_t + \beta_x^{(0)}\gamma_{t-x} \tag{15}$$

(referred to as model M in Haberman and Renshaw (2011) and M2 in Cairns et al. (2009), or more generally as the RH model). This model has also been

simplified to give

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x^{(1)}\kappa_t + \gamma_{t-x} \tag{16}$$

In this form, it is sometimes referred to as model H1 (in Haberman and Renshaw (2009, 2011) and Hunt and Villegas (2015) for instance). An extension of this and the LC2 model

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \gamma_{t-x} \tag{17}$$

was proposed in van Berkum et al. (2014) as part of a wider discussion of mortality models. More general extensions of the model of Wilmoth (1990) were also proposed in Hatzopoulos and Haberman (2011), which took the form

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + \sum_{i=2}^{N} \beta_x^{(i)}\kappa_t^{(i)} + \gamma_{t-x}^{(1)} + \sum_{j=2}^{M} \tilde{\beta}_x^{(j)}\gamma_{t-x}^{(j)} \tag{18}$$

However, in practice, the use of multiple cohort terms was found to be unnecessary and made the models difficult to fit to data.

A more ambitious attempt to hybridise features from the classic APC, LC and CBD models was performed in Plat (2009). This took the advantages of using a static age function from the LC and classic APC models, the use of multiple parametric age functions in the CBD models and the cohort term from the classic APC model to give

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + (x-\bar{x})^+\kappa_t^{(3)} + \gamma_{t-x} \tag{19}$$

where $z^+ \equiv max(z, 0)$. This "Plat model" was also simplified in Plat (2009) by removing the third age/period term to give the "reduced Plat model"

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \gamma_{t-x} \tag{20}$$

It was proposed that this simpler form of the model was more suitable for modelling mortality at high ages. Other variations of this were proposed in O'Hare and Li (2012), with the emphasis on finding a model suitable for modelling mortality at low ages

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \left((\bar{x}-x)^+\right)^2\kappa_t^{(3)} + \gamma_{t-x} \tag{21}$$

and in Börger et al. (2013)

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + (x - x_h)^+\kappa_t^{(3)} + (x_l - x)^+\kappa_t^{(4)} + \gamma_{t-x} \tag{22}$$

where the third and fourth terms are designed to capture changes in mortality at high ages (above fixed age x_h) and low ages (below fixed age x_l), respectively.

In response to this proliferation of models, Hunt and Blake (2014) introduced a "general procedure" for constructing bespoke mortality models within the class of APC models which are tailored to the specific features of the data in question. This procedure uses a "toolkit" of potential parametric forms for the age functions, the parameters of which are estimated from the data rather than fixed in advance. Models constructed by the general procedure tend to have many different forms of the age/period terms, much like the models of Plat (2009), O'Hare and Li (2012) and Börger et al. (2013), but the procedure gives a method for determining whether an additional term is justified by the data and what form that term should take.

As can be seen from this discussion, a wide variety of APC mortality models have been proposed in recent years. These differ substantially in their choice of the link function, $\eta_{x,t}$, nature of the age functions, either $\beta_x^{(i)}$ or $f^{(i)}(x)$, and whether or not to include a cohort term. We therefore seek to investigate the advantages and disadvantages of each of these choices in the following sections in order to better understand how to select a model appropriate to the needs of the model user.

4 Response variable and link function

When studying mortality, we typically assume that members of the population of interest experience the same instantaneous hazard rate of mortality $\mu_{x,t}$ at age x and time t (also called the "force of mortality"). In practice, however, observed data is usually grouped into discrete age and period bands and therefore modelling mortality is usually conducted using discrete time models.

In order to use the continuous force of mortality in a discrete age/period setting, it is commonly assumed that mortality rates do not change within each age and period band. Mathematically, this means that $\mu_{x,t}$ is assumed to be constant within ages and within years:

$$\mu_{x,t} = \mu_{\lfloor x \rfloor, \lfloor t \rfloor} \tag{23}$$

where $\lfloor x \rfloor$ is the integer part of x. This assumption is generally reasonable for most ages of interest (typically under age 100). Above this age, the populations under observation and, correspondingly, the number of deaths tend to be quite low, which means that the practical impact of this assumption breaking down is quite small over most ages. With the assumption that the force of mortality is constant over each age/period band, we can calculate the probability of survival for an individual of age x over the year t, $p_{x,t} = 1 - q_{x,t}$, as

$$\begin{aligned} p_{x,t} &= \exp\left(-\int_0^1 \mu_{x+s,t+s} ds\right) \\ &= \exp\left(-\mu_{x,t}\right) \end{aligned} \tag{24}$$

and the central mortality rate, $m_{x,t}$, as

$$\begin{aligned} m_{x,t} &= \frac{\int_0^1 \mu_{x+s,t+s} \, \exp\left(-\int_0^s \mu_{x+u,t+u} du\right) ds}{\int_0^1 \exp\left(-\int_0^s \mu_{x+u,t+u} du\right) ds} \\ &= \mu_{x,t} \end{aligned} \tag{25}$$

Almost all APC mortality models either use $\mu_{x,t}$ (or equivalently $m_{x,t}$) or $q_{x,t}$ as the response variable for mortality.

These two choices for the response variable reflect the two models for the random number of deaths, $D_{x,t}$, widely used in demography and actuarial science. Under the binomial assumption, the expected number of deaths is given by $\mathbb{E}(D_{x,t}) = E^0_{x,t} q_{x,t}$, the initial number of people alive (or initial exposure to risk) multiplied by the probability of death over the year. The probability of death can therefore be estimated as the observed number of deaths divided by the initial exposure to risk, $\hat{q}_{x,t} = \frac{d_{x,t}}{E^0_{x,t}}$.[3] Under the Poisson assumption, the expected number of deaths is given by $\mathbb{E}(D_{x,t}) = E^c_{x,t} m_{x,t}$,

[3]Where $d_{x,t}$ is the observation of the random death count, $D_{x,t}$.

i.e., the central exposure to risk (the average number of people alive which is used as a proxy for the total number of person-years lived) multiplied by the central mortality rate, $\hat{m}_{x,t} = \frac{d_{x,t}}{E^c_{x,t}}$.

This leads to the conclusion that the model for the response variable should be motivated by the format of the available data. The use of the Poisson model requires central exposures to risk which are widely available, for instance from the Human Mortality Database.[4] The use of the binomial model requires initial exposures to risk which are less commonly available for large populations (though may be more available for smaller populations) but can be approximated from the central exposures.

Asymptotically, for large populations and low death rates, the two approaches give similar results. It has been argued[5] that the binomial approach gives transformed mortality rates which are closer to being linear at the highest ages. However, it is at these ages that our assumption of a constant force of mortality within ages and years in Equation 23 starts to break down. This violates the core assumption underpinning the discrete time approach, making comparisons between models at high ages somewhat spurious.[6] Arguments on these lines should therefore be treated with caution.

In the Poisson and binomial models, the variance of the observations is also specified along with the mean. In practice, however, observations typically exhibit a greater variation than is predicted under either distribution - a phenomenon known as over-dispersion. One way of dealing with this is by fitting the model using the quasi-Poisson or quasi-binomial distributions, which add additional parameters to account for the over-dispersion. An alternative is to use the negative binomial model for death counts, as in Delwarde et al. (2007b), Renshaw and Haberman (2008) and Li et al. (2009). These approaches do not change the model structure in Equation 1, merely how it is fit to data. However, over-dispersion (along with significant correlation patterns within the fitted residuals) may also be a sign that the predictor structure is poorly chosen and so could be dealt with by selecting a more

[4]Human Mortality Database (2014).

[5]For instance, in Cairns et al. (2006a).

[6]One solution to this might be to assume a constant force of mortality over shorter age and period bands, for instance across months as in Gavrilov and Gavrilova (2011). However, data limitations at high ages tend to make this infeasible.

suitable alternative.

The link function, $\eta_{x,t}$, provides the connection between the observed data and the assume predictor structure. In the generalised linear model framework, there are several requirements which should be met for a good choice of link function. One of these is that the data should be transformed to obtain an approximately linear predictor structure (as opposed to, say, a multiplicative structure). Early static and dynamic mortality models used this as the sole requirement for the choice of $\eta_{x,t}$, which resulted in a range of choices being made, such as $\eta_x = \frac{q_x}{1-q_x}$ in Heligman and Pollard (1980), $\eta_{x,t} = \ln\left(\frac{q_{x,t}}{1-0.5q_{x,t}}\right)$ in Wilmoth (1990) and $\eta_{x,t} = \ln(\mu_{x,t})$ in Lee and Carter (1992). These models were then fitted using least squares estimation methods.

Least squares methods, however, do not account for the underlying distribution for $D_{x,t}$ and assume that the variance of observations is independent of the underlying exposures. However, this is not usually valid - observations are typically more variable at ages with low populations, such as those at high ages. More sophisticated methods of estimation, based on maximising the likelihood (Brouhns et al. (2002)) or, equivalently, minimising the scaled deviance (Renshaw and Haberman (2003a)), allow for this directly by making explicit reference to the underlying probability distribution of $D_{x,t}$. Although a number of potential link functions might be considered for either distribution of death counts (for instance, see Currie (2014)), practical considerations motivate using the canonical link function of the distribution $D_{x,t}$. The choice of this function also ensures that the fitted values of the response variable lie within the required range.[7] For a Poisson model of the death count, the canonical choice for the link function, $\eta_{x,t}$, is

$$\begin{aligned} \eta_{x,t} &= \ln(\mu_{x,t}) \\ \mathbb{E}[D_{x,t}] &= E^c_{x,t} e^{\eta_{x,t}} \\ \mathbb{V}ar(D_{x,t}) &= E^c_{x,t} e^{\eta_{x,t}} \end{aligned} \tag{26}$$

[7] i.e., $\mu_{x,t} \geq 0$ or $q_{x,t} \in (0,1)$.

whilst for the binomial model it is

$$\begin{aligned} \eta_{x,t} &= \text{logit}(q_{x,t}) \\ \mathbb{E}[D_{x,t}] &= E^0_{x,t} \frac{e^{\eta_{x,t}}}{1+e^{\eta_{x,t}}} \\ \mathbb{V}ar(D_{x,t}) &= E^0_{x,t} \frac{e^{\eta_{x,t}}}{(1+e^{\eta_{x,t}})^2} \end{aligned} \tag{27}$$

Using the canonical link function also has the desirable property that it simplifies estimation by maximum likelihood or minimal deviance considerably easier. For Poisson death counts using the log link function, the likelihood function is

$$\mathcal{L} = \sum_{x,t} W_{x,t} \left(d_{x,t} \ln(E^c_{x,t}\mu_{x,t}) - E^c_{x,t}\mu_{x,t} - \ln(d_{x,t}!) \right) \tag{28}$$

whilst for binomial death counts and the logit link function, the likelihood function is

$$\begin{aligned} \mathcal{L} = \sum_{x,t} W_{x,t} \big(& d_{x,t} \ln(q_{x,t}) + (E^0_{x,t} - d_{x,t}) \ln(1-q_{x,t}) \\ & + \ln(E^0_{x,t}!) - \ln((E^0_{x,t} - d_{x,t})!) - \ln(d_{x,t}!) \big) \end{aligned} \tag{29}$$

where $W_{x,t}$ are $\{0, 1\}$ weights. When using Newton-Raphson techniques to maximise the likelihood, we need to calculate the first and second derivatives of the log-likelihood function with respect to the parameters (e.g., see

Brouhns et al. (2002)), the forms of which are

$$\begin{aligned}
\frac{d\mathcal{L}}{d\alpha_x} &= \sum_t \left(d_{x,t} - \mathbb{E}[D_{x,t}]\right) \\
\frac{d^2\mathcal{L}}{d\left(\alpha_x\right)^2} &= -\sum_t \mathbb{V}ar(D_{x,t}) \\
\frac{d\mathcal{L}}{d\beta_x^{(i)}} &= \sum_t \left(d_{x,t} - \mathbb{E}[D_{x,t}]\right)\kappa_t^{(i)} \\
\frac{d^2\mathcal{L}}{d\left(\beta_x^{(i)}\right)^2} &= -\sum_t \mathbb{V}ar(D_{x,t})\left(\kappa_t^{(i)}\right)^2 \\
\frac{d\mathcal{L}}{d\kappa_t^{(i)}} &= \sum_x \left(d_{x,t} - \mathbb{E}[D_{x,t}]\right)\beta_x^{(i)} \\
\frac{d^2\mathcal{L}}{d\left(\kappa_t^{(i)}\right)^2} &= -\sum_x \mathbb{V}ar(D_{x,t})\left(\beta_x^{(i)}\right)^2 \\
\frac{d\mathcal{L}}{d\gamma_y} &= \sum_x \left(d_{x,x+y} - \mathbb{E}[D_{x,x+y}]\right)\beta_x^{(0)} \\
\frac{d^2\mathcal{L}}{d(\gamma_y)^2} &= -\sum_x \mathbb{V}ar(D_{x,x+y})\left(\beta_x^{(0)}\right)^2
\end{aligned}$$

These are simple to compute quickly if the canonical link is used. Alternative link structures require more complicated algorithms[8] which it may be desirable to avoid.

Any decisions regarding the choice of response variable and link function should take the following into account:

- The choice of probability distribution should reflect the available data - the binomial distribution is the natural choice with initial exposures to risk, whilst the Poisson distribution is more natural for model users with central exposures.

[8]See, for instance, the estimation of models in the CBD family using the LifeMetrics code in Coughlan et al. (2007) where a Poisson distribution of deaths is assumed with a logit link function.

- The choice of response variable follows naturally from the probability distribution - $\mu_{x,t}$ is the variable of interest in the Poisson distribution and $q_{x,t}$ in the binomial distribution.

- The appropriate canonical link function, $\eta_{x,t}$, follows naturally from the probability distribution selected. While other link functions can be chosen, such a choice would probably require further justification.

In practice, most modellers use the $\ln(\mu_{x,t})$ approach, i.e., a log link function, and assume the death count is a Poisson random variable. These models include those proposed in Brouhns et al. (2002), Renshaw and Haberman (2003b, 2006), Plat (2009), Haberman and Renshaw (2009) and O'Hare and Li (2012). However, the reasons for this are mainly historical, since they are based on the LC model where the log link function was chosen simply to obtain a linear predictor structure rather than with reference to the underlying distribution of the death counts or the available data. The alternative $\text{logit}(q_{x,t})$ approach has mainly been adopted by the CBD family of mortality models (Cairns et al. (2006a) and the extensions of this model in Cairns et al. (2009)),[9] and also in Aro and Pennanen (2011).

5 Static age function

A static age function, α_x, has been used in many mortality models from Hobcraft et al. (1982) and Lee and Carter (1992) onwards. By construction, this captures the features of the mortality curve across the age range of the data which do not change with time. A typical example of such a function, from the LC model fitted to male data from the USA (downloaded from the Human Mortality Database (2014)) for the period 1933 to 2007, is shown in Figure 1. Across the full age range, this shows features such as the excess number of deaths due to infant mortality at very low ages and accidents at young adult ages, which are common across both time periods and countries.

Some models, most notably those in the CBD family of mortality models and that in Aro and Pennanen (2011), dispense with the need for an explicit static age function by implicitly assuming that it can be approximated by a

[9] These models do not draw a direct link between the use of the logit function and binomial death counts. However, this connection is made explicit in Haberman and Renshaw (2011) and Currie (2014).

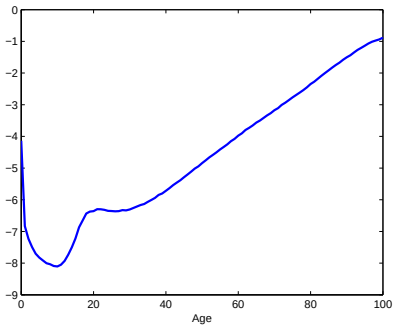

Figure 1: α_x static age function for the LC model fitted to US male data 1933-2007

simpler function of age and combining it into the age/period terms. To do this, the static age function needs to be a linear combination of the other age functions in the model, i.e.,

$$\alpha_x = \sum_{i=1}^{N} \alpha^{(i)} \beta_x^{(i)}$$

This is easiest to do when the age functions $\beta_x^{(i)}$ are known in advance of fitting the model to data. For example, the model of Cairns et al. (2006a) implicitly assumes that mortality rates are approximately linear at the ages of interest and therefore can be combined with the other terms in the model.

Doing this improves the parsimony of the model by reducing the number of free parameters considerably. However, it does so at the expense of limiting the model to only those parts of the age range where this assumption is approximately valid, typically at higher ages.

It also means that the age/period terms in the model do two tasks simultaneously: capturing the time-independent shape of mortality and describing the structure of the deviations from this shape. Including a static age function in the model therefore allows each term in the model to focus on doing one job optimally. The extent to which this is desirable will depend upon the modeller's preference for a parsimonious fit to historical data against the more detailed identification and projection of evolving trends.

6 Age/period terms

The age/period terms in an APC model typically capture the majority of the dynamic structure present in the underlying data. They consist of age functions, $\beta_x^{(i)}$, describing how the particular mortality effects are distributed across ages, which are multiplied by period functions, $\kappa_t^{(i)}$, which explain how they evolve with time.

One of the key distinctions between APC models is whether the age effects are modelled using "non-parametric" or "parametric" age functions. Some mortality models have age functions which are "non-parametric" in the sense that values of $\beta_x^{(i)}$ at different ages, x, are fitted without imposing any a priori structure. Age is treated as an unknown factor in the model rather than a regressor with a known structure.[10] Other mortality models have age functions which are "parametric", since they take a specific functional form that is defined by an algebraic formula.[11]

We should note that our definitions of the terms "non-parametric" and "parametric" differs from other definitions of these terms used in statistics and actuarial science. For the avoidance of doubt, we use the terms to specifically refer to the structure of the age/period terms, and they have no implication for the methods used to fit the model to data. For example, Haberman and Renshaw (2009, 2011) used the term "parametric" to refer to the predictor structure for general APC mortality models, and describe any models within this class as "parametric mortality models". Alternatively, "parametric" can refer to the underlying distributional assumptions for the model and the methods used to fit it to data as such, the assumption of a Poisson distribution of deaths and maximum likelihood estimation would lead to a "parametric mortality model" under this definition. Our usage of these terms is restricted solely to the form of the age effects.

[10] For this reason, we could alternatively refer to non-parametric age functions as "factorial" age functions.

[11] For this reason, these age functions could also be called "formulaic".

6.1 Non-parametric age functions

The models in the LC family, discussed in Section 3.1, all used non-parametric age functions. Such models are necessarily bilinear, as both age and period are unknown factors. Non-parametric approaches emerge naturally from model fitting techniques based principal component analysis (PCA), although they can easily be deployed in a generalised non-linear modelling or maximum likelihood framework.[12] The non-parametric approach also easily extends to an arbitrary number of age/period terms as discussed in Section 3.1. The number of age/period terms in the model is then selected with reference to the data, rather than having been prescribed in advance.

The main advantage of this approach is that the shapes of the age functions are chosen to maximise the fit to the data. This means that each term extracts the maximum amount of information from the data possible. For example, the terms produced by PCA are ranked in order of information extraction - as measured by the percentage of the total variability in the data explained - which makes it possible to select algorithmically an optimal number of terms in the model.

The non-parametric approach is also very flexible. It can be applied easily across a variety of datasets, as described, for example, in Tuljapurkar et al. (2000), which used the LC model to fit data from a number of developed nations. Similarly, the non-parametric approach can be used across the full age range, whilst parametric age functions are often only suitable for limited age ranges. It also avoids subjective judgements in constructing the model, as terms are fitted automatically to maximise the fit to data. This ability to objectively pick out the most important structure within the data is used as the starting point for the "general procedure" for constructing mortality models outlined in Hunt and Blake (2014).

However, non-parametric approaches have a number of downsides. Most importantly, the form of the non-parametric age functions generated usually

[12]PCA assumes homogenous, normally distributed residuals and, therefore, is inconsistent with the underlying binomial or Poisson distribution for the death count process. However, the estimates obtained for the parameters using PCA can be used as the starting point for methods such as maximum likelihood which use the death count process to allow for heterogeneity caused by differences in the underlying exposures.

lack demographic significance. To illustrate, Figure 2 shows the β_x age function produced by fitting the LC model to the same data for men in the US used in Section 5. It shows that, over the period, improvements in mortality rates have been far faster at young ages (below 20, but especially at age one) than at higher ages, where improvements have been more evenly distributed across ages. It is very difficult to conceive of an explanation for this shape which does not involve multiple drivers of changing mortality rates over the period - such as improved hygiene reducing mortality across all ages, childhood vaccination programmes reducing the number of deaths amongst the very young, and improved treatment of cardio-vascular disease in later life.

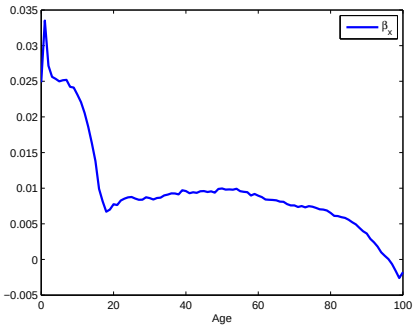

Figure 2: β_x age function for the LC model fitted to US male data 1933-2007

This has consequences when we fit and project the model. Drivers of mortality are combined into a single term if they are correlated over the historical period of the data (e.g., they go from a high level of mortality to a lower level over the period). However, these combinations may not be appropriate over subsets of the period range. For example, Carter and Prskawetz (2001) found that the form of β_x changes substantially if the LC model is fitted to different subintervals of the data, as different medical and socio-economic causes of mortality become more or less important.

Such combinations of drivers may also be inappropriate when we come to making forecasts using the model. For instance, we may believe that the shape of β_x in Figure 2 is due to a combination of childhood immunisation programmes and improved cardio-vascular care for the elderly. When projecting mortality, we may wish to allow the latter to continue to improve in

future. but believe that we are unlikely to see further reductions in mortality due to increased vaccination of children. Using a term which combines both these causes can lead to projections of mortality rates which do not appear to be plausible, e.g., when high rates of improvement in mortality are projected at ages where mortality rates are already very low.

Furthermore, the model does not require that the non-parametric forms are continuous.[13] This can lead to projections which have discontinuous mortality rates and so are not biologically reasonable[14] if projected far into the future. It is possible to smooth the non-parametric age functions to avoid this, as discussed in Delwarde et al. (2007a) or Hyndman and Ullah (2007). However, this complicates the structure of the model and introduces subjective decisions regarding the degree of smoothing which would need careful justification.

6.2 Parametric age functions

As discussed earlier, a parametric age function takes a specific functional form, i.e. $\beta_x = f(x)$. The classic APC model, the CBD family of models and many of the hybrid models discussed in Section 3.3 use parametric age functions in order to capture specific features of the evolution of mortality rates across the age range.

We can see that most of these models have a linear predictor structure, rather than possessing any bilinear terms where the age function also needs to be fitted to the data. This means that they are conventional generalised linear models and can be fitted using standard techniques. However, the use of parametric age functions does not necessarily imply linearity. For instance, consider the model

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \exp(-\lambda x)\kappa_t^{(2)}$$

[13] This can be seen with the sharp peak at β_1 in Figure 2.

[14] Introduced in Cairns et al. (2006b) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*". Note that biological reasonableness is a property of observable quantities such as life expectancies or mortality rates, in contrast to demographic significance which relates to our interpretation of the terms in a model.

Here, $f^{(2)}(x) = \exp(-\lambda x)$ is parametric in our sense of having a prescribed functional form, but λ can be a free parameter set with reference to the data and so the age/period term is bilinear and the model cannot be estimated via a generalised linear model. Age functions including free parameters are not widely used, as the higher-order age functions in the models of Plat (2009), Aro and Pennanen (2011), O'Hare and Li (2012) and Börger et al. (2013) have parameters which are set a priori. In principle, however, these models could be extended to allow these parameters to vary to find the value which fits the data best. In addition, many of the age functions used in the "general procedure" of Hunt and Blake (2014) possess free parameters and therefore are bilinear, parametric age/period terms.

One of the major advantages of using parametric age functions is that they reduce considerably the number of free parameters needing to be fitted for each age/period term, leading to more parsimonious models. This, in turn, means that more parameters can be devoted to detecting other features of interest within the data, such as additional structure across time and year of birth.

Furthermore, because the shapes of the age functions are known, each term can be assigned a specific demographic significance by the user. To illustrate, the first age/period term in the models of Equations 13 and 6 are constant across all ages. This can be explained in terms of specific phenomena which are universal across the age range (such as improved hygiene), in contrast with the shape seen in Figure 2. It will also allow trends which are correlated (such as improving levels of medical care for the elderly and the specific efforts to tackle childhood infectious diseases) to be given their own age/period terms with appropriate parametric age functions, which is impossible with a non-parametric approach.

However, this flexibility comes at a cost. Parametric age functions are often only suitable over limited age ranges. While this is an advantage in that it allows for greater interpretability of their demographic significance, it means that models with parametric age functions are often not suitable over the full age range. For instance, even if the CBD model were extended with a static age function, it is unlikely that the two age/period terms are sufficient to capture the variability of mortality rates at younger ages. In order to construct a model appropriate across the full age range, we would

have to add additional age/period terms to the model.

In addition, models with parametric age functions often give a poorer fit to the data compared to a model with the same number of non-parametric age/period terms, especially using measures of goodness of fit that do not (or only weakly) penalise the number of free parameters in the model. This is because the additional freedom in the non-parametric age function can be used to capture more of the structure in the data than if the form of the age function is prescribed at the outset.

These problems can be rectified, in part, through adding new terms to the model. However, we will need to decide on the appropriate form for these new terms, which can very often be difficult. One approach adopted for some of the extensions to the CBD model in Cairns et al. (2009) is to select age functions from the same family - in this case polynomials of increasing order. Alternatively, more exotic functions can be used, as in the models of Plat (2009) and O'Hare and Li (2012), but often there does not appear to be any underlying rationale for the selection made. In the end, expert judgement is needed to assess whether a new term added to the model genuinely represents the remaining unexplained trend in the data or merely reflects the expectation of the modeller as to what should be present.

7 Cohort effects

It is a widely held belief that the different life histories of individuals should lead to systematic difference between different cohorts of people (as summarised by their year of birth). These are often known as "cohort effects". As Hobcraft et al. (1982), Willets (1999) and Murphy (2009) discussed, the term "cohort effect" is largely descriptive, and some care needs to be taken in ascribing the causal factors specific to certain years of birth which might plausibly influence the mortality rate of a cohort across their entire life. We might, for instance, consider an epidemic which, in addition to raising mortality rates at the time it is raging, had a selective effect on the survival of infants. This might lead to systematic differences in mortality between those born during the epidemic and those born shortly before or afterwards. However, the evidence from natural experiments (summarised in Murphy (2009)) is equivocal, which means that the existence of true cohort effects is still

controversial to some extent, as discussed in Murphy (2010).

In practice, however, observed data from a number of countries appears to exhibit cohort features and so it is prudent to allow for these when modelling mortality. In the UK, apparent cohort effects have been identified in the general population (specifically in the work of Willets (1999, 2004), Continuous Mortality Investigation (2002) and Richards (2008)) and models allowing for cohort parameters outperformed those which did not in Cairns et al. (2009).

Our subjective demographic significance of a cohort effect is one which increases or reduces mortality at all ages for individuals born in a specific generation (typically lasting 10-15 years or less). To construct a mortality model, we need to translate this demographic significance into a set of properties we desire the parameters in our model to possess. More specifically, we can say that our intuition regarding the cohort effects implies that they should:

- be small relative to the effects of age and period;
- not have any systematic trends in their expected value or variability;
- have a mean across cohorts of zero (i.e., cohort effects should represent deviations from a typical hypothetical reference level);
- have some autocorrelation: it is reasonable to believe that cohorts born in successive years should experience similar life histories and so exhibit similar cohort effects, unless there happen to be exceptional circumstances facing a particular birth year;
- not exhibit indefinite persistence: the factors influencing the specific mortality of the generation born today should be essentially independent of the specific mortality of their grandparents, for example;
- ideally be mean reverting (as a consequence of the previous two points), as the specific events impacting one cohort wear off in subsequent years of birth; and
- be demographically significant, so we can relate features of a plot of cohort effects to specific socio-economic and medical influences on the population.

In a well-specified mortality model, many of these properties emerge naturally from the fitted parameters. Some, such as the level of the mean of the cohort parameters, can be imposed via identifiability constraints, which change the values of the cohort parameters but not the fit of the model to data. However, this is not always the case, and we may sometimes have to discard some of our intuitive properties based on the evidence of the model. For instance, we can see that in Plat (2009), the historical cohort parameters have a clear trend and may be non-stationary.

We would also like estimates of our cohort parameters to be robust, both across different models and when comparing them with the residuals from the corresponding age/period mortality model, as in Wilmoth (1990). For instance, the plots of cohort parameters for the same datasets in Cairns et al. (2009) show that the features identified are not robust between different models, which weakens any demographic significance we place on them. However, there are a number of practical problems that makes finding cohort parameter estimates that are robust and well specified a harder task than the estimation of age and period parameters.

First, because age, period and cohort are linearly dependent ($y + x = t$), we cannot treat them in isolation of each other.[15] Wilmoth (1990) argued that it is impossible to apportion objectively low frequency (slowly varying) temporal dependence in mortality data between age/period and cohort effects. We therefore are forced to make a subjective choice to give primacy to two of the relevant dimensions. Because we naturally observe cross sections of mortality rates across ages in different calendar years, the data will naturally form a rectangular age/period grid. This means that the natural choice is to give primacy to age and period effects and to try to explain as much of the structure in the data with reference to these dimensions as possible before consideration of effects across cohorts.[16]

This then leads to the conclusion that if the cohort effects are to be taken as of secondary importance, the structure in the model included to cap-

[15] We also suffer from the problem that the parameters in the model may not be uniquely identified. This topic and its implications for forecasting are discussed further in Hunt and Blake (2015).

[16] See Alai and Sherris (2012) for an example of a model which gives primacy to cohort parameters.

ture them should be as simple as possible. Indeed, some have argued that cohort effects might not exist at all and are merely the result of poorly specified age/period effects.[17] A model user operating under such a belief would therefore omit any age/cohort terms from the model entirely. A high standard of evidence for the inclusion of an age/cohort term is therefore desirable.

If an age/cohort term is to be included and if age/cohort interactions are taken to be of secondary importance, the desire for parsimony in the cohort terms leads to two further conclusions which have been adopted by the majority of model users. First, most models only include one cohort term on the grounds that it is hard to believe and to demonstrate that one generation could experience two different independent lifelong effects. Nevertheless, the model proposed in Hatzopoulos and Haberman (2011) allows for multiple cohort effects.

Second, many models set $\beta_x^{(0)} = 1$, leading to a more parsimonious model. This restriction allows the cohort parameters to represent consistently higher or lower mortality rates across all ages, which accords with our demographic interpretation of cohort effects. In particular, while a cohort effect which is stronger at some ages than others does not seem unreasonable in principle, the notion of a cohort effect that increases mortality rates at some ages but decreases them at others conflicts with our interpretation of the demographic significance of a cohort effect. This situation is possible with a non-parametric form for $\beta_x^{(0)}$ unless it is artificially constrained to be greater than zero. In addition, issues have also been reported concerning the robustness of fitting models such as that of Renshaw and Haberman (2006) with a non-parametric $\beta_x^{(0)}$ term, for instance by Continuous Mortality Investigation (2006) and Cairns et al. (2009).[18] However, this problem is not universal and a linear parametric form for $\beta_x^{(0)}$ was proposed in model M8 in Equation 11 by Cairns et al. (2009) and has been found to be robust and to fit the data well in van Berkum et al. (2014).

Cohort parameters also present specific problems in estimation which

[17] For instance, Cairns et al. (2011) raised *"the possibility that cohort effects might be partially or completely replaced by well-chosen age and period effects"* and also see Murphy (2010).

[18] See Hunt and Villegas (2015) for a discussion and potential solution for this issue.

again suggests that a parsimonious model structure be used when including them. Because we naturally observe cross sections of mortality rates across ages in different calendar years, we will have a limited numbers of observations for the earliest and latest birth cohorts. This makes estimates of these cohort parameters more uncertain. For instance, the last observed year of birth will only have one observation for it, which can therefore be fit perfectly by the cohort term. This is undesirable and so, in practice, many modellers do not estimate cohort parameters for a number of the earliest and latest years of birth in the data (for instance in Renshaw and Haberman (2006) and Cairns et al. (2009)).

Related to this is the fact that the observations for early and late years of birth will only cover a subset of the age range. For instance, the most recent cohorts will only have observations for the youngest ages. Any misspecification of age/period terms affecting these ages will therefore bias the estimation of these cohort parameters. This is especially important for the most recent cohorts, for which we will only have a small number of observations on their early-age mortality where most mortality models have the greatest difficulty modelling the age/period patterns of mortality and where there will be relatively few deaths. Any poorly specified age/period terms at these ages will therefore lead to structure in the data being wrongly attributed to the cohort effect for the most recent years of birth.

As an example of this, there are specific biological factors which lead to mortality in the first year of life evolving differently from mortality rates at subsequent ages. This effect is best captured through an age/period interaction. In a poorly specified age/period mortality model, this cannot be captured adequately, leading to large residuals when fitting mortality rates at this age. Adding a cohort term to such a model will mean that the fitting procedure will try to use the extra parameters to "solve" this problem and so will bias the cohort parameters in order to "fix" what is genuinely an age/period issue. This bias will get more pronounced for more recent years of birth, where observations of the first year of life form an increasing proportion of the total observations for each new cohort.

In models which give primacy to age/period effects, it is therefore important to ensure that the age/period structure is fully specified before an age/cohort term is added. When forecasting mortality rates, it is of great

practical importance that the cohort parameters in an APC model are well specified and estimated robustly. As cohort effects represent lifelong mortality effects, mis-specifications of the cohort parameters at low ages will bias forecasts for these cohorts as they age.

In summary, the inclusion of a cohort term in a mortality model presents the user with a number of important issues which need to be addressed. In some cases, the model user may consider that cohort effects are not significant and prefer a model which does not include them. However, in other populations, there is evidence to support their inclusion. In such cases, it is necessary to ensure that the age/period structure in the model is well specified and able to capture the majority of structure in the data. A simple and parsimonious cohort term can then be included to capture the effects of year of birth in a way that avoids the cohort term attempting to capture residual age/period structure.

8 Classification of APC mortality models

Despite the recent rapid proliferation in the number of mortality models proposed, the majority of mortality models in discrete time are part of the same underlying APC family. This then leads to the natural question of how mortality models can be classified. One such classification of APC mortality models, based on the models' "ancestry", was given in Cairns (2014) and described in Section 3. However, as can be seen in Section 3.3, this is less helpful when considering the new range of hybrid models. Unlike species of animal, however, the ancestry of mortality models is not unique, since they can adopt features from different pre-existing mortality models to achieve different aims.

Hence, we believe that a sensible classification should pay more attention to the terms within the model and how they achieve the aims of the model user. To address these, we must ask a number of questions when constructing an APC mortality model, but especially the following:

- What response variable and link function should we use?
- Should we include an explicit static age function?

- Should we use parametric or non-parametric age functions? If so, how many age/period terms should we use?
- Should we include a cohort term? If so, should it be modified across the age range by a $\beta_x^{(0)}$ age function?

What we offer below is a simple classification of mortality models, based on what we consider to be the most important differences in structure between them, as determined by the answers to these questions.

We believe that the first two questions above are straightforward. The modeller's choice for the response variable should depend on the data available to them rather than on any more fundamental consideration. This, in turn, leads to a natural choice for the link function, namely, the canonical link function for the chosen distribution of deaths. Whilst it is possible to use combinations of response variable and link function other than the natural choices, there is often no good reason to do this and practical reasons discussed in Section 4 why it should be avoided.

Second, it can be argued that all mortality models use a static age function; it is just that models such as the CBD model of Cairns et al. (2006a) use it implicitly with a distinct parametric structure that enables it to be combined with other terms on the model. Such a choice may be desirable for models limited to specific sections of the age range where the parametric structure is appropriate in order to obtain greater parsimony. However, it does not change anything fundamental about the model.

We are then left with the two more substantive questions - the choice between parametric and non-parametric age functions and the inclusion of a cohort term. Both of these reflect fundamental differences in approach which lead to important mathematical and qualitative differences between the models. Historically, however, cohort parameters have often been seen as an optional addition to a pre-existing mortality model, especially because the age/period terms are usually given primacy due to the reasons discussed in Section 7. We, therefore, believe that the most important division amongst APC models to be between the use of parametric and non-parametric age functions.

The optimum number of age/period terms will then depend on the nature of the age functions chosen to define these terms. In models with non-parametric age functions, it is relatively simple to add additional age/period terms and optimise their number based on a goodness of fit criteria. In models with parametric age functions, however, the number of age functions needs to be defined a priori along with their functional form. If new terms are to be added to an existing model, it is a non-trivial task to select an appropriate form for them. To solve this problem, Hunt and Blake (2014) introduced a "general procedure" to both select the form of the parametric age functions and determine an optimum number of age/period terms in a new mortality model.

Based on this analysis, we propose a simple classification of mortality models in Figure 3. Obviously this classification is not exhaustive, as new models and variations of existing models are continuously being proposed. It is also not unique, since a different ordering of the questions asked when constructing a mortality model would yield a different family tree. However, we have found it to be a useful framework when considering the selection of an existing mortality model or when constructing a new one (such as in Hunt and Blake (2014)).

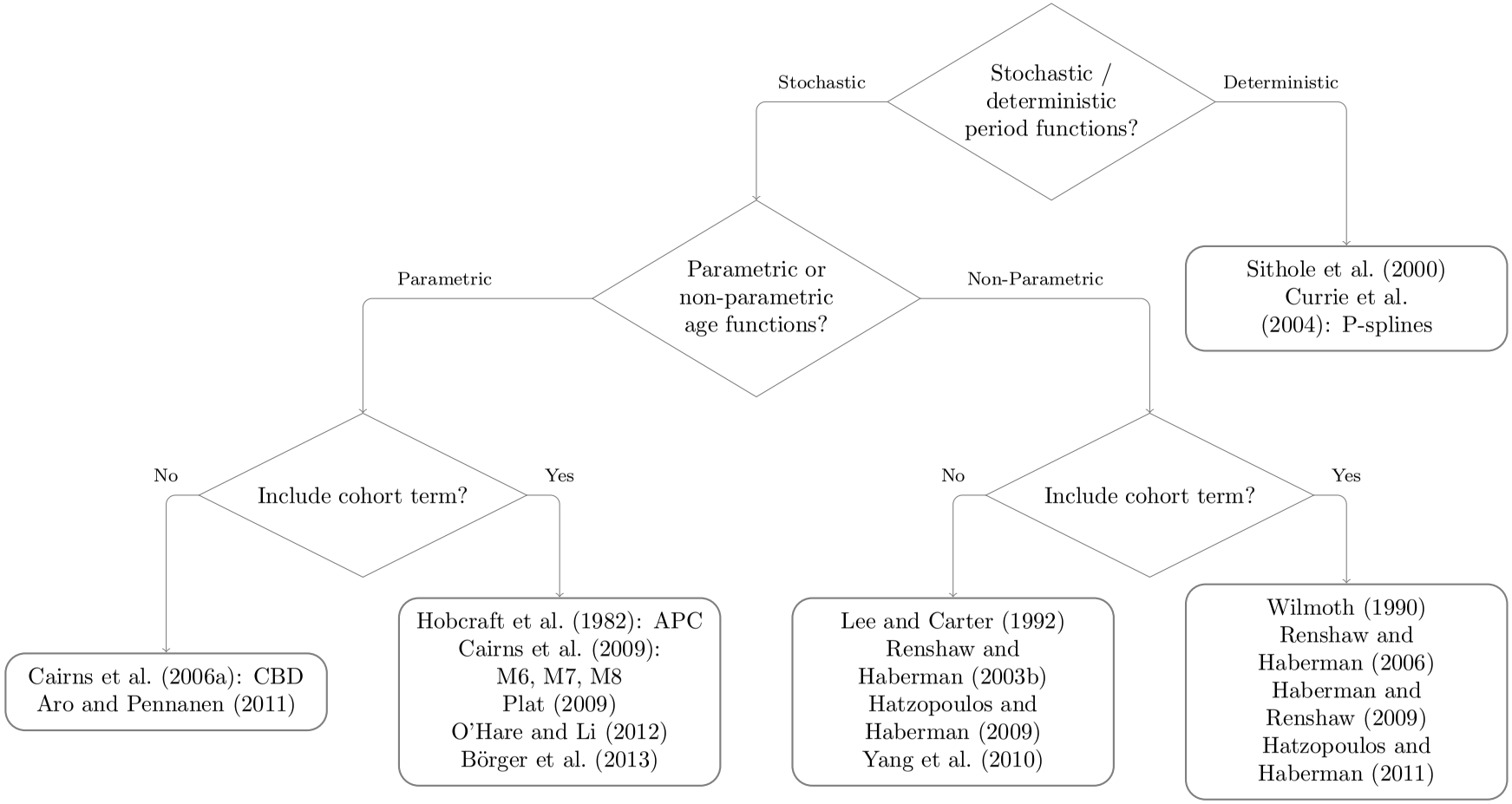

Figure 3: A simple classification of mortality models

9 Conclusions

The increasing number of age/period/cohort models being used to study and project mortality rates has made a general consideration of the APC structure necessary. We have reviewed the evolution of this class of models in terms of the "ancestry" of the mortality models which have been proposed to date. However, we need to go beyond a classification based on ancestry in order to understand the properties of specific mortality models and their advantages and disadvantages. A systematic and complete understanding of this structure allows us to either select or construct the most appropriate model for the dataset and the purpose. We have set out five principles which need to be considered before an APC mortality model can be used or constructed:

1. The response variable being modelled should match the data available. The link functions should follow naturally from the nature of the response variable, e.g., a Poisson distribution for the number of deaths should lead naturally to a log-link function.

2. A static age function should generally be included and made explicit in the model. If a parametric structure is assumed for the static age function, this should be made explicit and the limitations this places on the age range over which the model is suitable should be made clear.

3. The user should justify the choice of a non-parametric or parametric structure for the age functions. Both are appropriate in different circumstances. However, the user of a model should be explicit in the trade-offs they are making between goodness of fit and demographic significance.

4. The use of a cohort term is usually desirable to capture structure across year of birth in the data. However, such a term can be omitted if the evidence does not support its inclusion.

5. When cohort terms are included in a mortality model, they should be made as simple as possible in order to give robust parameter estimates. This will often lead to using a single cohort term and setting $\beta_x^{(0)} = 1$.

We therefore believe that the examination of the structure of APC mortality models in this paper has direct practical application when using and

developing these models and enables a natural classification to be developed. A proper understanding of the models can therefore help practitioners analyse how mortality has evolved in the past and how it may evolve in future, which is of great importance for the financial and social management of longevity risk.

References

Alai, D. H., Sherris, M., 2012. Rethinking age-period-cohort mortality trend models. Scandinavian Actuarial Journal 18 (2), 452–466.

Aro, H., Pennanen, T., 2011. A user-friendly approach to stochastic mortality modelling. European Actuarial Journal 1 (S2), 151–167.

Baran, S., Gáll, J., Ispány, M., Pap, G., 2007. Forecasting Hungarian mortality rates using the Lee-Carter methods. Acta Oeconomica 57 (1), 1–20.

Booth, H., Maindonald, J., Smith, L., 2002. Applying Lee-Carter under conditions of variable mortality decline. Population Studies 56 (3), 325–36.

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. M., Vermunt, J., 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. Insurance: Mathematics and Economics 31 (3), 373–393.

Cairns, A. J., 2014. Modelling and management of longevity risk. Tech. rep., Heriot-Watt University, Edinburgh.

Cairns, A. J., Blake, D., Dowd, K., 2006a. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. Journal of Risk and Insurance 73 (4), 687–718.

Cairns, A. J., Blake, D., Dowd, K., 2006b. Pricing death: Frameworks for the valuation and securitization of mortality risk. ASTIN Bulletin 36 (1), 79–120.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., 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, A. J., Blake, D., Dowd, K., Coughlan, G. D., 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.

Callot, L., Haldrup, N., Lamb, M. K., 2014. Deterministic and stochastic trends in the Lee-Carter mortality model. Tech. rep., Aarhus University.

Carstensen, B., 2007. Age-period-cohort models for the Lexis diagram. Statistics in Medicine 26, 3018–3045.

Carter, L. R., Prskawetz, A., 2001. Examining structural shifts in mortality using the Lee-Carter method. Tech. rep., Max Planck Institute for Demographic Research.

Clayton, D., Schifflers, E., 1987. Models for temporal variation in cancer rates. II: Age-period-cohort models. Statistics in Medicine 6 (4), 469–81.

Continuous Mortality Investigation, 2002. Working Paper 1 - An interim basis for adjusting the "92" series mortality projections for cohort effects. URL `http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-1`

Continuous Mortality Investigation, 2006. Working Paper 20 - Stochastic projection methodologies: Further progress and P-spline model features, example results and implications.
URL `http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-20`

Coughlan, G. D., Epstein, D., Ong, A., Sinha, A., 2007. LifeMetrics: A toolkit for measuring and managing longevity and mortality risks. Technical document. JPMorgan Pension Advisory Group.

Currie, I. D., 2006. Smoothing and forecasting mortality rates with P-splines.
URL `http://www.ma.hw.ac.uk/~iain/research/talks/Mortality.pdf`

Currie, I. D., 2014. On fitting generalized linear and non-linear models of mortality. Scandinavian Actuarial Journal, Forthcoming.

Currie, I. D., Durbán, M., Eilers, P., 2004. Smoothing and forecasting mortality rates. Statistical Modelling 4 (4), 279–298.

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

Delwarde, A., Denuit, M. M., Partrat, C., 2007b. Negative binomial version of the Lee-Carter model for mortality forecasting. Applied Stochastic Models in Business and Industry 23 (5), 385–401.

Fienberg, S. E., Mason, W. M., 1979. Identification and estimation of age-period-cohort models in the analysis of discrete archival data. Sociological Methodology 10, 1–67.

Gavrilov, L. A., Gavrilova, N. S., 2011. Mortality measurement at advanced ages: A study of the social security administration death master file. North American Actuarial Journal 15 (3), 432–447.

Glenn, N., 1976. Cohort analysts' futile quest: Statistical attempts to separate age, period and cohort effects. American Sociological Review 41 (5), 900–904.

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.

Haberman, S., Renshaw, A., 2012. Parametric mortality improvement rate modelling and projecting. Insurance: Mathematics and Economics 50 (3), 309–333.

Haberman, S., Renshaw, A., 2013. Modelling and projecting mortality improvement rates using a cohort perspective. Insurance: Mathematics and Economics 53 (1), 150–168.

Hanewald, K., 2011. Explaining mortality dynamics: The role of macroeconomic fluctuations and cause of deaths trends. North American Actuarial Journal 15 (2), 290–314.

Hatzopoulos, P., Haberman, S., 2009. A parameterized approach to modeling and forecasting mortality. Insurance: Mathematics and Economics 44 (1), 103–123.

Hatzopoulos, P., Haberman, S., 2011. A dynamic parameterization modeling for the age-period-cohort mortality. Insurance: Mathematics and Economics 49 (2), 155–174.

Heligman, L., Pollard, J., 1980. The age pattern of mortality. Journal of the Institute of Actuaries 107 (1), 49–80.

Hobcraft, J., Menken, J., Preston, S. H., 1982. Age, period and cohort effects in demography: A review. Population Index 48 (1), 4–43.

Holford, T. R., 1983. The estimation of age, period and cohort effects for vital rates. Biometrics 39 (2), 311–24.

Human Mortality Database, 2014. Human Mortality Database. Tech. rep., University of California, Berkeley and Max Planck Institute for Demographic Research.
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. Identifiability in age/period/cohort mortality models. Tech. rep., Pensions Institute PI 15-09, Cass Business School, City University London.

Hunt, A., Villegas, A. M., 2015. Robustness and convergence in the Lee-Carter model with cohort effects. Insurance: Mathematics and Economics 64, 186–202.

Hyndman, R., Ullah, M., 2007. Robust forecasting of mortality and fertility rates: A functional data approach. Computational Statistics & Data Analysis 51 (10), 4942–4956.

Kleinow, T., Cairns, A. J., 2013. Mortality and smoking prevalence: An empirical investigation in ten developed countries. British Actuarial Journal 18 (2), 452–466.

Kuang, D., Nielsen, B., Nielsen, J. P., 2008a. Forecasting with the age-period-cohort model and the extended chain-ladder model. Biometrika 95 (4), 987–991.

Kuang, D., Nielsen, B., Nielsen, J. P., 2008b. Identification of the age-period-cohort model and the extended chain-ladder model. Biometrika 95 (4), 979–986.

Lee, R. D., 2000. The Lee-Carter method for forecasting mortality, with various extensions and applications. North American Actuarial Journal 4 (1), 80–93.

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

Lee, R. D., Miller, T., 2001. Evaluating the performance of the Lee-Carter method for forecasting mortality. Demography 38 (4), 537–549.

Li, J. S.-H., Hardy, M. R., Tan, K. S., 2009. Uncertainty in mortality forecasting: An extension to the classical Lee-Carter approach. ASTIN Bulletin 39 (1), 137–164.

Mitchell, D., Brockett, P. L., Mendoza-Arriaga, R., Muthuraman, K., 2013. Modeling and forecasting mortality rates. Insurance: Mathematics and Economics 52 (2), 275–285.

Murphy, M., 2009. The "golden generations" in historical context. British Actuarial Journal 15 (S1), 151–184.

Murphy, M., 2010. Re-examining the dominance of birth cohort effects on mortality. Population and Development Review 36 (2), 365–90.

O'Brien, R. M., 2000. Age period cohort characteristic models. Social Science Research 29 (1), 123–139.

O'Brien, R. M., 2011. Constrained estimators and age-period-cohort models. Sociological Methods & Research 40 (3), 419–452.

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), 124–9.

Plat, R., 2009. On stochastic mortality modeling. Insurance: Mathematics and Economics 45 (3), 393–404.

Reichmuth, W., Sarferaz, S., 2008. Bayesian demographic modeling and forecasting: An application to U.S. mortality. Tech. rep., Humbolt University, Berlin.

Renshaw, A., Haberman, S., 2003a. Lee-Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections. Journal of the Royal Statistical Society: Series C (Applied Statistics) 52 (1), 119–137.

Renshaw, A., Haberman, S., 2003b. Lee-Carter mortality forecasting with age-specific enhancement. Insurance: Mathematics and Economics 33 (2), 255–272.

Renshaw, A., Haberman, S., 2003c. On the forecasting of mortality reduction factors. Insurance: Mathematics and Economics 32 (3), 379–401.

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.

Richards, S. J., 2008. Detecting year-of-birth mortality patterns with limited data. Journal of the Royal Statistical Society: Series A (Statistics in Society) 171 (1), 279–298.

Rodgers, W., 1982. Estimable functions of age, period, and cohort effects. American Sociological Review 47 (6), 774–787.

Sithole, T., Haberman, S., Verrall, R., 2000. An investigation into parametric models for mortality projections, with applications to immediate annuitants' and life office pensioners' data. Insurance: Mathematics and Economics 27 (3), 285–312.

Tuljapurkar, S., Li, N., Boe, C., 2000. A universal pattern of mortality decline in the G7 countries. Nature 405 (6788), 789–792.

van Berkum, F., Antonio, K., Vellekoop, M. H., 2014. The impact of multiple structural changes on mortality predictions. Scandinavian Actuarial Journal, Forthcoming.

Villegas, A. M., Millossovich, P., Kaishev, V. K., 2015. StMoMo: An R Package for Stochastic Mortality Modelling.
URL URLhttp://CRAN.R-project.org/package=StMoMo

Wang, H., Preston, S. H., 2009. Forecasting United States mortality using cohort smoking histories. Proceedings of the National Academy of Sciences of the United States of America 106 (2), 393–8.

Wang, J. L., Huang, H., Yang, S. S., Tsai, J. T., 2009. An optimal product mix for hedging longevity risk in life insurance companies: The immunization theory approach. Journal of Risk and Insurance 77 (2), 473–497.

Willets, R., 1999. Mortality in the next millennium. Staple Inn Actuarial Society.

Willets, R., 2004. The cohort effect: Insights and explanations. British Actuarial Journal 10 (4), 833–877.

Wilmoth, J. R., 1990. Variation in vital rates by age, period and cohort. Sociological Methodology 20, 295–335.

Wilmoth, J. R., 1993. Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Tech. rep., Department of Demography, University of California, Berkeley.

Yang, B., Li, J., Balasooriya, U., 2015. Using bootstrapping to incorporate model error for risk-neutral pricing of longevity risk. Insurance: Mathematics and Economics 62, 16–27.

Yang, S. S., Yue, J. C., Huang, H., 2010. Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models. Insurance: Mathematics and Economics 46 (1), 254–270.

Yang, Y., Fu, W. J., Land, K. C., 2004. A methodological comparison of age-period-cohort models: The intrinsic estimator and conventional generalized linear models. Sociological Methodology 34, 75–110.