



DISCUSSION PAPER PI-1205

Modelling the Cohort Effect in CBD Models
Using a Piecewise Linear Approach

Y. Zhao and P.J. Sweeting

May 2012

ISSN 1367-580X

The Pensions Institute
Cass Business School
City University London
106 Bunhill Row
London EC1Y 8TZ
UNITED KINGDOM

<http://www.pensions-institute.org/>

Modelling the Cohort Effect in CBD Models Using a Piecewise Linear Approach

*Y. Zhao (University of Kent) and
P.J. Sweeting (University of Kent; J.P. Morgan Asset Management)*

Abstract

This paper discusses a new pattern of mortality model which is built on the form and knowledge of the two-factor mortality model named after its designers Cairns, Blake and Dowd (2006). This model – the CBD model – is widely used and has been extended by the authors in a number of ways, including by the use of a cohort effect. In this paper, we propose a range of new parsimonious approaches to model the cohort effect. Instead of adding a cohort factor to an age-period model we model the effect by building discontinuities into the pattern of rates within each year. The fit of the resulting models is close to that available from the best of the CBD derivatives.

Keywords

Mortality Risk; Stochastic Modelling; Kink points; Cairns-Blake-Dowd; Cohort effects; Linear Functions

Contact addresses

Y. Zhao, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent, CT2 7NZ. E-mail: yz76@kent.ac.uk

P. J. Sweeting, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent, CT2 7NZ. E-mail: p.j.sweeting@kent.ac.uk

1. Introduction

In the last twenty years, great improvements in mortality modelling have taken place, from the Lee-Carter model (Lee and Carter 1992) and its extensions (Brouhns, et al 2002; Renshaw & Haberman 2003, 2006; Continuous Mortality Investigation Bureau 2005) to the CBD model (Cairns, et al 2006b) and its extensions (Cairns, et al 2009). Cairns et al (2009) comprehensively compare and rank these models in terms of the Bayesian Information Criterion (BIC), described by Herman (2006). According to the conclusion from Cairns et al, it is evident that the cohort factor representing the effect of the birth year improves the fit of mortality models considerably.

In general, there are three main extrapolative models in use. The first is the Lee-Carter – or L-C model family. The L-C model, which is a single-factor model, assumes no smoothness across ages or years; it was the first stochastic extrapolative model to be developed. The second is the P-splines model which uses splines to impose smoothness across years and ages (this is not considered). The last is CBD model family which is a two-factor model (with or without cohort effects) which assumes smoothness across ages in same year, but makes no assumption between smoothness in different years.

Figure 1 reveals the main cohort effect in terms of the cohort parameter from the Renshaw Haberman model, described below. Noteworthy discontinuities occur corresponding to the ending of hostilities in the First and Second World Wars. It is possible to identify the first of these with the 1919 influenza epidemic. There is another discontinuity in 1887, which can be traced to a set of outlying observations. This is possibly due to mis-stated exposures for this particular cohort (Renshaw, et al 2006). In addition to the discontinuities, there is also a notable change in mortality improvement rates for certain cohort. Willets (2003) highlights rapid mortality improvements in generations born between 1925 and 1945. In his paper, the cohort effect is explored in other fields such as epidemiology. This gives a greater understanding of the cohort effect, in which the cause of death by prevalence of smoking from one generation to the next, heart disease and breast cancer are cited as major factors.

Before introducing how the cohort effect is allowed for in different models, let us revisit the L-C model. The L-C model (Lee & Carter 1992) can be expressed in the following way:

$$\log m(t, x) = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \varepsilon_{tx} \quad (\text{Equation 1})$$

where, $\beta_x^{(1)}$ is the age factor; $\beta_x^{(2)} \kappa_t^{(2)}$ is period factor; ε_{tx} refers to an error term.

In the model of Renshaw & Haberman (2006) – referred to here as the R-H model – the L-C model is improved by adding a cohort factor $\beta_x^{(3)} \gamma_{t-x}^{(3)}$, i.e.:

$$\log m(t, x) = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)} + \varepsilon_{tx} \quad (\text{Equation 2})$$

where, $\beta_x^{(1)}$ is the age factor; $\beta_x^{(2)} \kappa_t^{(2)}$ is period factor; ε_{tx} refers to an error term. While in the cohort factor $\beta_x^{(3)} \gamma_{t-x}^{(3)}$, $\gamma_{t-x}^{(3)}$ is the pure cohort factor for year of birth $t - x$, and $\beta_x^{(3)}$ is an age weight.

To highlight the cohort effect, we fit the R-H model to data from 1961 to 2009, and ages from 0 to 89. In Figure 1 the trend of cohort parameter in the R-H model shows the impact of the cohort effect. As mentioned in Willets (2003) and Renshaw and Haberman (2006), several noticeable features are obvious and should be taken into account, the most noteworthy of which is the steep relative fall in mortality rates for those born between the mid 1920's and the early 1940's, but it is also interesting to note the discontinuities in $\gamma_{t-x}^{(3)}$ in and around 1919 and 1945 – in other words, around the ends of the two world wars. The

timing of these two discontinuities, combined with the fact each anomaly has a fall below the trend followed by a rise above it, that suggests that the end of the world wars coincided with data anomalies. This is important, as it suggests that any parameterisations suggesting changes around these points should be viewed with suspicion. It also implies that any smoothing of these anomalies should not have a significant effect on expectations of life or annuity functions.

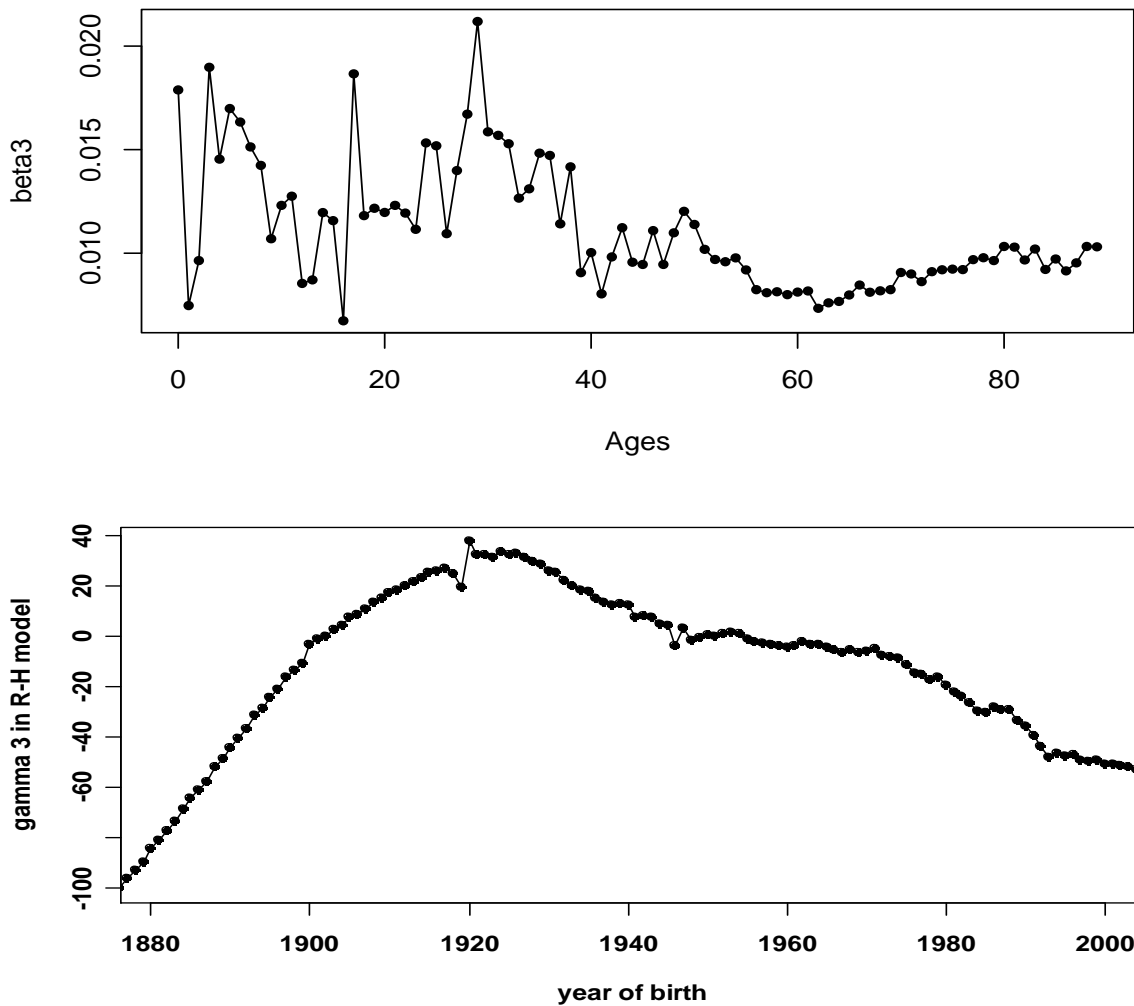


Figure 1: R-H model, England & Wales, male, ages 0-89, period 1961-2009

The CBD model is given as the following (Cairns, et al 2006b):

$$\text{logit } q(t, x) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \varepsilon_{tx} \quad (\text{Equation 3})$$

where $q(t, x)$ is the initial mortality rate for a life aged x in year t ; \bar{x} is the average of the ages x ; $\kappa_t^{(1)}$ denotes the intercept parameter in year t ; $\kappa_t^{(2)}$ denotes the slope parameter in year t ; and ε_{tx} refers to an error term.

In Cairns et al (2009), the CBD model is improved by adding a cohort factor $\gamma_{t-x}^{(3)}$. The improved model (named as M6 in Cairns et al 2009) is:

$$\text{logit } q(t, x) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_{t-x}^{(3)} + \varepsilon_{tx} \quad (\text{Equation 4})$$

When using data from 1961 to 2004 and ages from 60 to 89, in Figure 2, we get the intercept parameter $\kappa_t^{(1)}$ and the slope parameter $\kappa_t^{(2)}$ as plotted in the first row, while the trend of cohort parameter $\gamma_{t-x}^{(3)}$ in the second row shows the impact of the cohort effect, reiterating the conclusions of Willets (2003) and Renshaw and Haberman (2006).

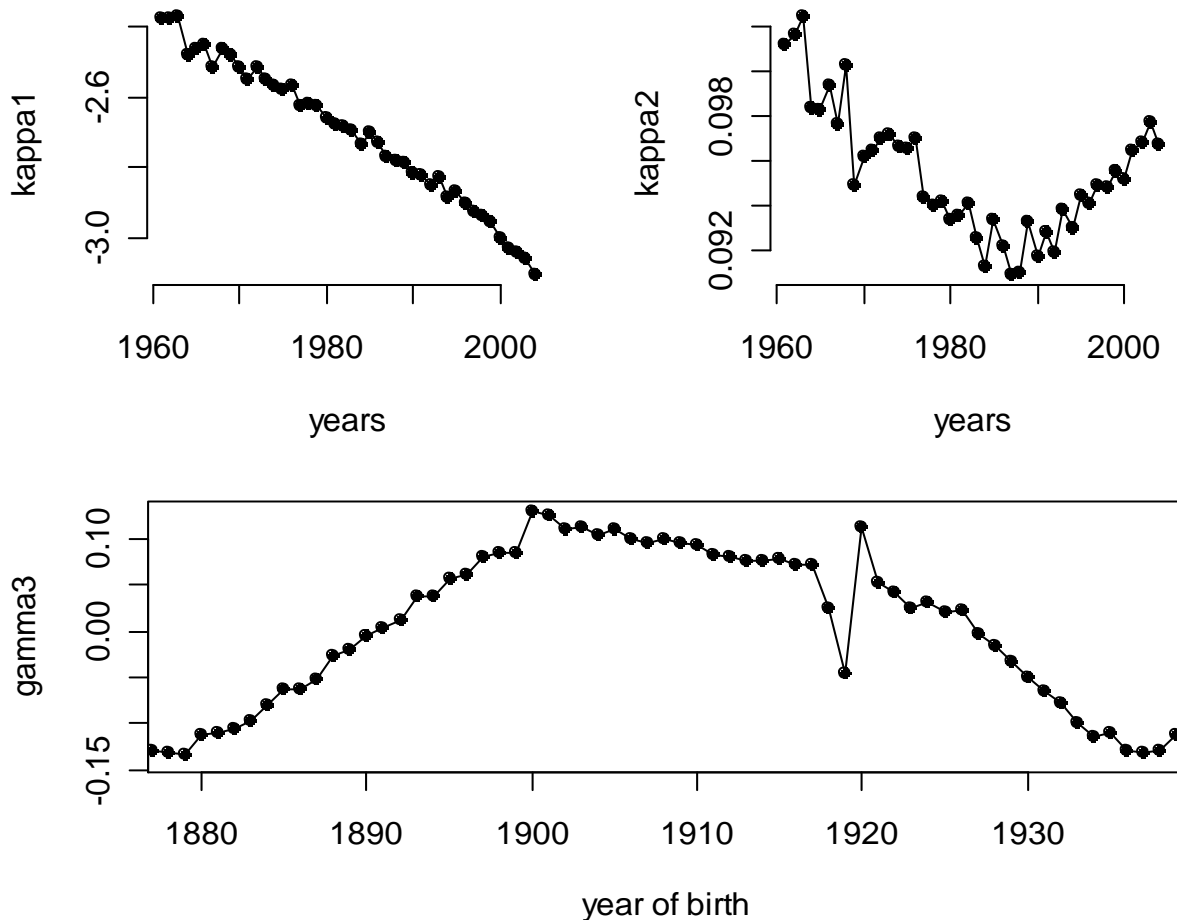


Figure 2: M6, England & Wales, male, ages 60-89, period 1961-2004

From the biological explanation, the trends of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ in Figure 2 give the general trends in which kappa1 implies that mortality has been improving over years, and $\kappa_t^{(2)}$ indicates that mortality rates at higher ages are decreasing at a lower rate (Cairns et al. 2006). The new models in this paper still follow this biological pattern.

2. Proposed Models

2.1 Motivation

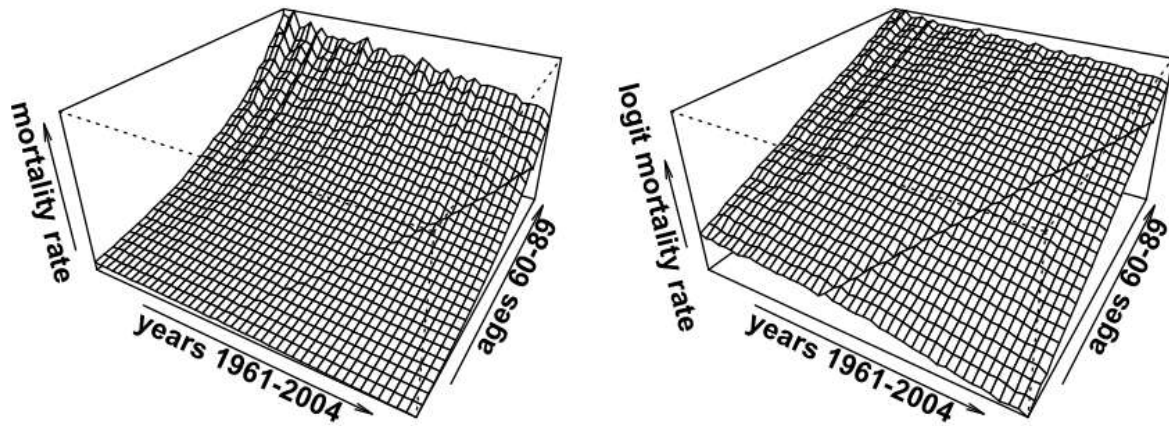


Figure 3: England & Wales, ages 60-89, period 1961-2004

When we plot the mortality rate and logit of mortality rate in Figure 3, England and Wales ages from 60-89, from 1961-2004, it can be seen that mortality rates have been gradually falling over the period of observation. Furthermore, the cohort effects can be easily seen as diagonal features.

In each year, mortality rates can be fitted as a smooth linear function using the CBD model. However, because of the cohort effect, there are clearly diagonal breaks across the entire surface, as shown in Figure 3. As discussed above, model M6 deals with this by adding a separate cohort factor to the logit of mortality rates for each year of birth. However, an alternative is to use two or more linear functions to fit $\text{logit}(q_{tx})$ in each calendar year. To do this, we need to know where the breaks occur exactly in each year and whether a break is statistically significant.

For instance, Figure 4 illustrates $\text{logit}(q_x)$ in 1985, 1990, 1995 and 2000 respectively. It is obvious that inflections exist around age 65, 70, 75 and 80 respectively, which also means these people were born around 1920. Hence, a hypothesis that the significant kink point in each year is caused by the birth year close to 1920 can be held. The method and the relative models to deal with this problem are shown later in this paper.

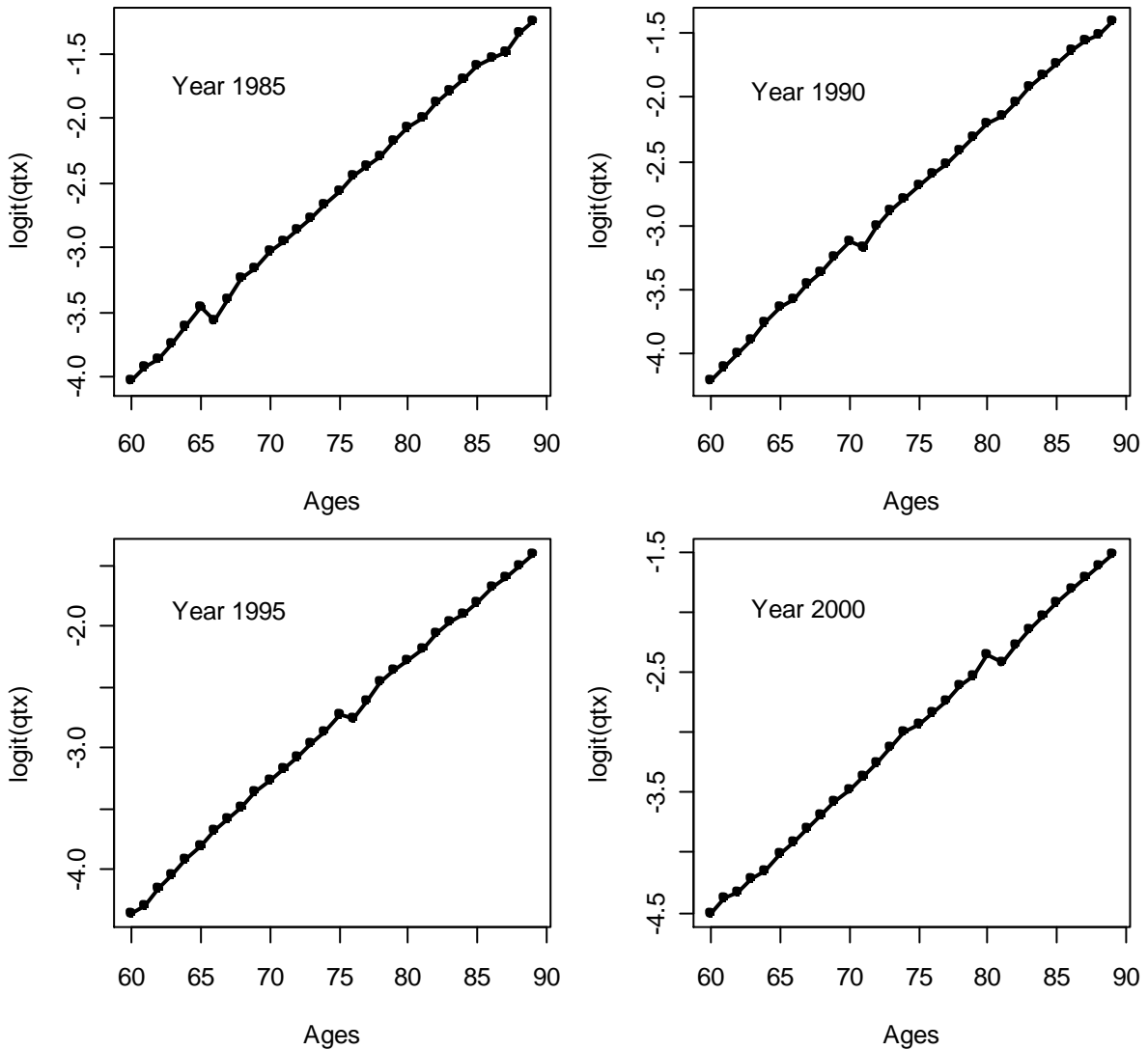


Figure 4: examples

In Sweeting (2011), it is clear that there are breaks in the trend for both $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ over long time horizons. The author used a Chow test (Chow 1960) to examine structural breaks. The same approach can be used to investigate structural breaks in the mortality rates. In this paper, after adding the kink points which are examined under the method of maximum likelihood, the New Models (NMs) replace the linear function of logit mortality used in the CBD model in each year, with a varying number of linear functions instead, the number of functions representing the number of cohorts.

2.2 Motivational representation

Suppose there is a single kink in the logit mortality rate in a particular year t occurring for lives born before or after year c . This is shown stylistically for some year t for lives between ages 60 and 89 in Figure 5. As can be seen, two linear functions can be used to fit the logit of mortality rates, instead of the single linear function of the CBD model.

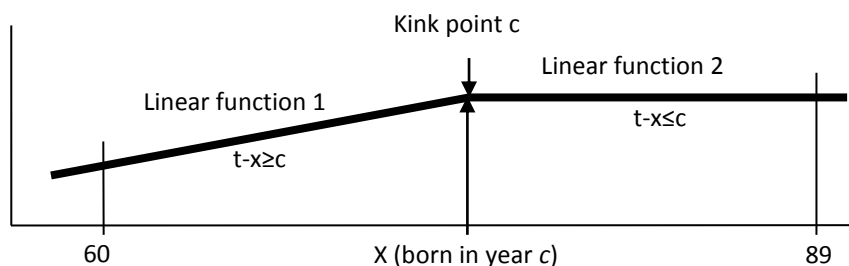


Figure 5: logit of mortality rate in some year t

This model (denoted as New Model One, or NM1) can be described as:

$$\text{logit}(q_{tx}) = \begin{cases} \kappa_t^{11} + \kappa_t^{12}(x - \bar{x}) + \varepsilon_{tx}, & t - x \geq c \\ \kappa_t^{21} + \kappa_t^{22}(x - \bar{x}) + \varepsilon_{tx}, & t - x \leq c \end{cases} \quad (\text{Equation 5})$$

$$\text{Constraint: } \kappa_t^{11} + \kappa_t^{12}(t - c - \bar{x}) = \kappa_t^{21} + \kappa_t^{22}(t - c - \bar{x})$$

where year c is a kink point of birth year; $\kappa_t^{i1}, i = 1, 2$ denotes intercept factor in year t : $i = 1$, fits the equation for ages less than or equal to $t + c$, while $i = 2$ fits the equation to those whose ages are greater than or equal to $t - c$; $\kappa_t^{i2}, i = 1, 2$ denotes slope factor in year t , with i having the same meaning as before; and ε_{tx} expresses an error term. The model is fitted using the Poisson maximum likelihood approach described by Brouhns et al (2002). The constraint ensures that there is no first order discontinuity between the two linear functions.

An important feature of this model is that the term c remains constant across the years investigated. Once a value for c has been found that maximises the log likelihood (introduced in section 3.2), the cohort birth year is fixed at year $t - c$: it does not change through the use of the Genetic Algorithm Method (GAM), introduced later.

New Model Two (NM2) allows for two kink points rather than one, which improves the goodness-of-fit. The mechanism using two kink points is similar to NM1, breaking the original smoothly linear function of CBD model into a piecewise linear function with the two kink points for given birth years as shown in Figure 6. This shows an instance when both kink points are located between age 60 and 89, although the kink points can also be outside the span of ages shown here.

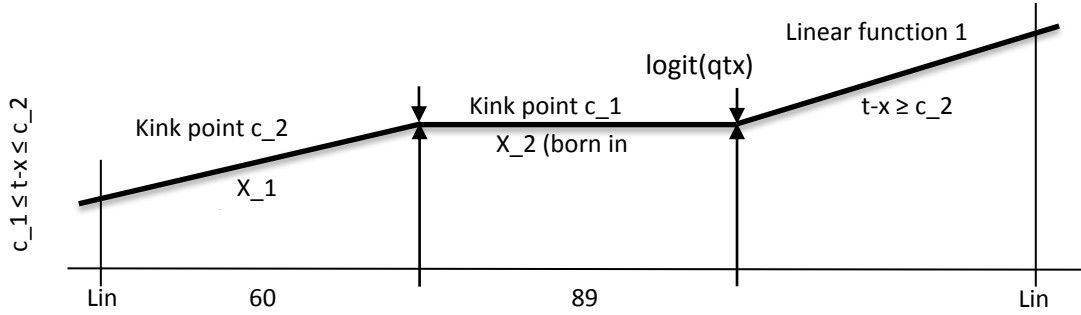


Figure 6: logit of mortality rate in some year t

This model is also fitted under the assumption that the number of deaths follows a Poisson distribution.

$$\logit(q_{tx}) = \begin{cases} \kappa_t^{11} + \kappa_t^{12}(x - \bar{x}) + \varepsilon_{tx}, & t - x \geq c_2 \\ \kappa_t^{21} + \kappa_t^{22}(x - \bar{x}) + \varepsilon_{tx}, & c_1 \leq t - x \leq c_2 \\ \kappa_t^{31} + \kappa_t^{32}(x - \bar{x}) + \varepsilon_{tx}, & t - x \leq c_1 \end{cases} \quad (\text{Equation 6})$$

Constraints:

$$\kappa_t^{11} + \kappa_t^{12}(t - c_2 - \bar{x}) = \kappa_t^{21} + \kappa_t^{22}(t - c_2 - \bar{x})$$

$$\kappa_t^{21} + \kappa_t^{22}(t - c_1 - \bar{x}) = \kappa_t^{31} + \kappa_t^{32}(t - c_1 - \bar{x})$$

where, years c_1, c_2 are the two kink points, and $c_1 < c_2$; $\kappa_t^{i1}, i = 1, 2, 3$ denotes intercept factor in year t : $i = 1$ fits the equation for ages less than and equal to $t - c_2$, $i = 2$ fits the equation for ages greater than or equal to $t - c_2$ but less than or equal to $t - c_1$, and $i = 3$ fits the equation for ages greater than or equal to $t - c_1$; $\kappa_t^{i2}, i = 1, 2, 3$ denotes slope factors for each year t , with i having the same meaning as before; ε_{tx} is an error term.

2.3 Alternative formulations

By the same reasoning, more kink points can be added. Therefore, a comprehensive model can be given by:

$$\logit(q_{tx}) = \begin{cases} \kappa_t^{11} + \kappa_t^{12}(x - \bar{x}) + \varepsilon_{tx}, & t - x \geq c_n \\ \kappa_t^{21} + \kappa_t^{22}(x - \bar{x}) + \varepsilon_{tx}, & c_{n-1} \leq t - x \leq c_n \\ \dots & \dots \\ \kappa_t^{(n+1)1} + \kappa_t^{(n+1)2}(x - \bar{x}) + \varepsilon_{tx}, & t - x \leq c_1 \end{cases} \quad (\text{Equation 7})$$

Constraints:

$$\kappa_t^{11} + \kappa_t^{12}(t - c_n - \bar{x}) = \kappa_t^{21} + \kappa_t^{22}(t - c_n - \bar{x})$$

$$\kappa_t^{n1} + \kappa_t^{n2}(t - c_1 - \bar{x}) = \overset{\dots}{\kappa_t^{(n+1)1}} + \kappa_t^{(n+1)2}(t - c_1 - \bar{x})$$

where the kink points are $c_i, i = 1, 2, \dots, n; c_1 < c_2 < \dots < c_{n-1} < c_n$.

3. Evaluation

3.1 Data

Cairns, et al (2009), use a consistent set of data to compare eight models. To enable comparison with those results, this paper uses the same data, namely mortality rates for England and Wales, males, age from 60 to 89 and year between 1961 and 2004, supported by the Human Mortality Database (HMD).

3.2 Estimation

Let the exposure in year t for age x be E_{tx} and let the central mortality rate be m_{tx} , where $t = 1961, 1962, \dots, 2004; x = 60, 61, \dots, 89$. Following a Poisson distribution, the number of deaths d_{tx} has mean $E_{tx} \times m_{tx}$, i.e. $d_{tx} \sim Poi(E_{tx}m_{tx})$. As mentioned earlier, this method is broadly adopted in the literature on mortality modelling, being described in detail by Brouhns et al. (2002).

Transformation of the mortality rates greatly aids modelling, with the log and logit of mortality rates being of primary interest (Booth & Tickle 2008). The CBD model employs q_{tx} , the initial mortality rate at time t for age x , as do the New Models which are extensions of CBD model in this paper. The approximate relationship between initial and central mortality rate, m_{tx} , is: $q_{tx} = 1 - e^{-m_{tx}}$, i.e. $m_{tx} = -\log(1 - q_{tx})$.

Based on the relationship above, the likelihood, L , can be given as:

$$L = \sum_{t,x} d_{tx} \log(E_{tx}m_{tx}) - E_{tx}m_{tx} - \log(d_{tx}!),$$

and the different models can be estimated by maximum likelihood (Cairns, et al 2009).

Under the assumption that the linear functions are not piecewise continuous, we can fit the observations separately after choosing possible kink points of cohort year and calculating the total maximum log likelihood. By comparing these values of maximum log likelihood from all different possibilities of kink points, we can find out the optimal positions for these points, which can be considered as the turning points of the improvement of mortality rate by cohort. The weakness of this method is that it only allows for the goodness of fit partially and separately rather than completely. Another visual reflect of the weakness is that there are gaps between linear functions existing at the kink points (see *Figure 7*).

In order to make the weakness “gaps” more transparent, NM4 is taken for example. NM4 is

$$\text{logit}(q_{tx}) = \begin{cases} \kappa_t^{11} + \kappa_t^{12}(x - \bar{x}) + \varepsilon_{tx}, & t - x \geq c_4 \\ \kappa_t^{21} + \kappa_t^{22}(x - \bar{x}) + \varepsilon_{tx}, & c_3 \leq t - x \leq c_4 \\ \kappa_t^{31} + \kappa_t^{32}(x - \bar{x}) + \varepsilon_{tx}, & c_2 \leq t - x \leq c_3 \\ \kappa_t^{41} + \kappa_t^{42}(x - \bar{x}) + \varepsilon_{tx}, & c_1 \leq t - x \leq c_2 \\ \kappa_t^{51} + \kappa_t^{52}(x - \bar{x}) + \varepsilon_{tx}, & t - x \leq c_1 \end{cases} \quad (\text{Equation 8})$$

where the kink points are $c_i, i = 1, 2, 3, 4; c_1 < c_2 < c_3 < c_4$.

So if we choose year 1985 as the example, the mortality rates against ages are shown in Figure 7. As can be seen, if the constraints are ignored, the gaps between adjacent linear functions are obvious, especially for the third kink for cohort year 1919.

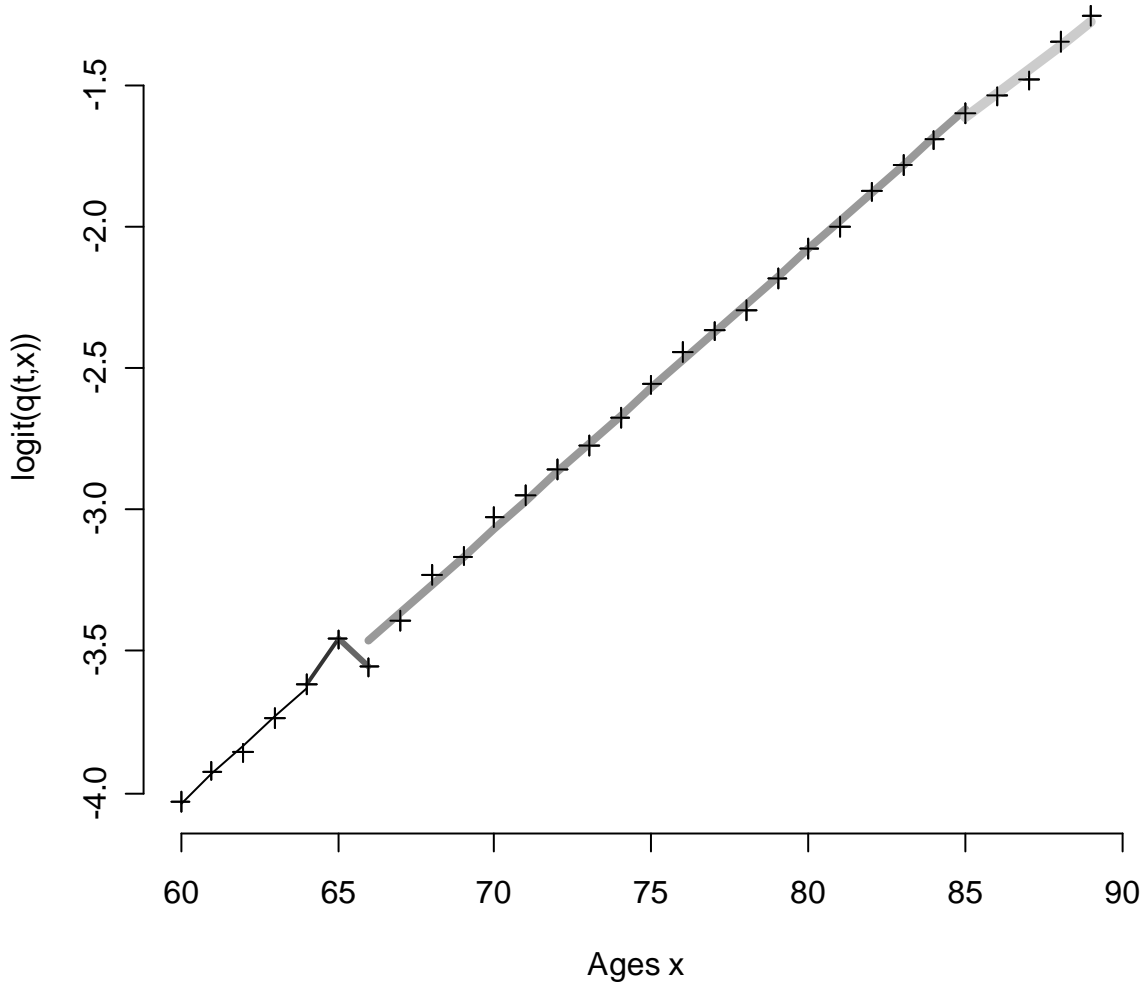


Figure 7: In year 1985

In practise, we believe that the changing of mortality rate follows a smoothly continuous process, even if it is not continuously differentiable. Hence, we need to find out a method which must enable the two adjacent linear functions connect to each other. The genetic algorithm method is a good way to deal with the problem.

3.3 Genetic Algorithm Method

The genetic algorithm method (GAM) is used to generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover (Bajpai et al. 2008).

First pioneered by John Holland in the 1960s, genetic algorithms (GAs) have been widely studied, experimented with and applied in many fields in engineering worlds. Not only do GAs provide an alternative method to solving problem, they consistently outperform other traditional methods in most problems. Many real world problems involved finding optimal parameters, which might prove difficult for traditional methods but ideal for GAs.

In the paper, we have “gap” problems between two linear functions, which are not consistent with reality. Considering the best fit, we have to adjust all the functions simultaneously. Not only must the adjacent functions be connected, but also the log likelihood must be at a maximum. Therefore, if in a given year, the mortality rates are fitted by k linear functions, we get $2k$ parameters (k slope factors and k intercept factors). Considering the constraints (*Equations 5 to 7*) then we have $k + 1$ parameters to be adjusted simultaneously in this year. The parameters are that slopes of each functions plus one intercept parameter of one selected function. The other intercept parameters can be calculated according to the constraints given in Equation 5 – 7. In a word, GAM helps us optimize problems in multi-dimensional environments.

Still taking NM4 for example, in year 1985, the mortality rates against ages are shown in Figure 8.

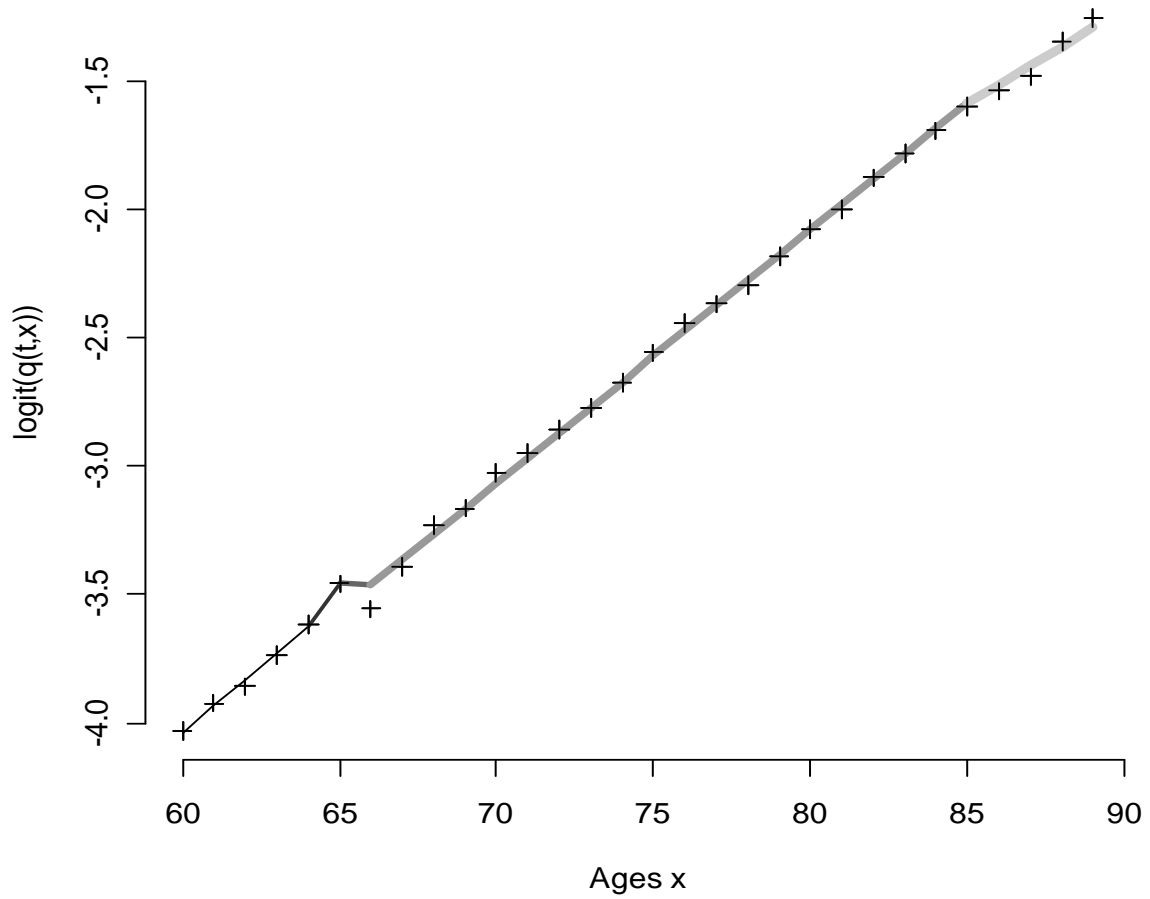


Figure 8: In year 1985

With help from GAM, it is obvious that in Figure 8 the “gaps” disappear, resulting in a more realistic outcome. And what is more, by using the GAM, the number of parameters actually is reduced from $2k$ to $k + 1$.

3.3 Results for Different Models

3.3.1 New Model One (NM1)

New Model One (see Formula NM1) only allows for one kink parameter c .

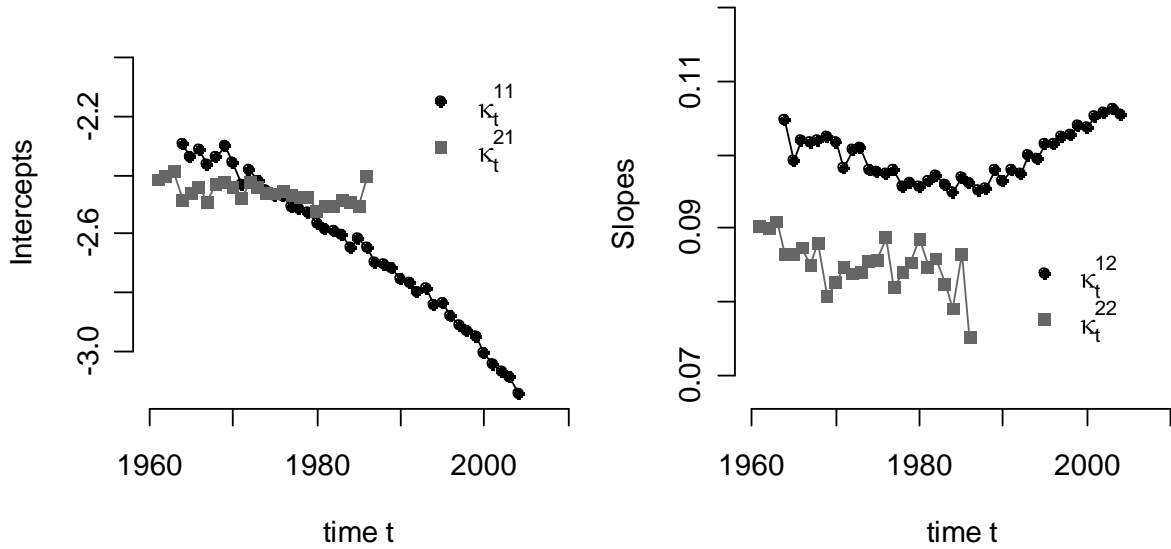


Figure 9: Parameters in NM1

For NM1, with the help of log likelihood maximisation and genetic algorithm methods, we find that the most likely position for the kink occurs at year of birth 1900. Figure 9 shows the results of estimated parameters. The black lines represent the statistical outcomes under the condition that $t - x \geq 1900$, i.e. the birth year is after or in year 1900; while the grey lines stand for the statistical outcomes under the condition that $t - x \leq 1900$, i.e. birth year is before or in year 1900.

The left graph in Figure 9 illustrates the trends of intercept parameters κ_t^{11} and κ_t^{21} with black and grey lines respectively. Because of limited data for some periods and cohorts, the values of some parameters fluctuate severely. For example, in NM1 the kink point is birth year 1900. This means that in 1962, people born in or after 1900 were less than or equal to age 62. When we fit the logit of mortality rates in 1962, there are only 3 data points that can be fitted by one linear function, since we are using only ages 60 and over. Hence, these parameters estimated using fewer than 4 data points are ignored. As illustrated, after ignoring these values, the trends of κ_t^{11} and κ_t^{21} show approximately linear functions with slight negative slopes. This makes each series amenable to projection using a random walk with drift. Similarly, the right graph in Figure 9 displays the trends of slope parameters, κ_t^{12} (black), κ_t^{22} (grey) respectively. Instead of the seemingly linear function for $\kappa_t^{(2)}$ in CBD model, the grey line rises after initially falling.

The CBD model projects factors κ_t^1 and κ_t^2 as a two-dimensional random walk with drift (Cairns et al 2008). For NM1, it therefore makes sense to project each of the pairs $\kappa_t^{11}, \kappa_t^{12}$ and $\kappa_t^{21}, \kappa_t^{22}$ in the same way: the data for κ_t^{11} and κ_t^{12} both relate to people born in or after 1900, whilst κ_t^{21} and κ_t^{22} both relate to people born up to and including 1900. However, virtually everyone born in or before 1900 has died so κ_t^{21} and κ_t^{22} have no effect on forecasting. In other words, κ_t^{11} and κ_t^{12} are of more interest to us. What is more, the influence of κ_t^{21} and κ_t^{22} on the analysis has been removed, improving the fit for κ_t^{11} and κ_t^{12} .

Figure 10 shows how the kink point works in a particular year, say 1975. As shown, the logit of initial mortality rates in year 1975 broken into two parts. The slim black line ends at age 75, and the thick grey line begins from 75. In other words, the kink point of birth year is exactly 1900 (1975-75).

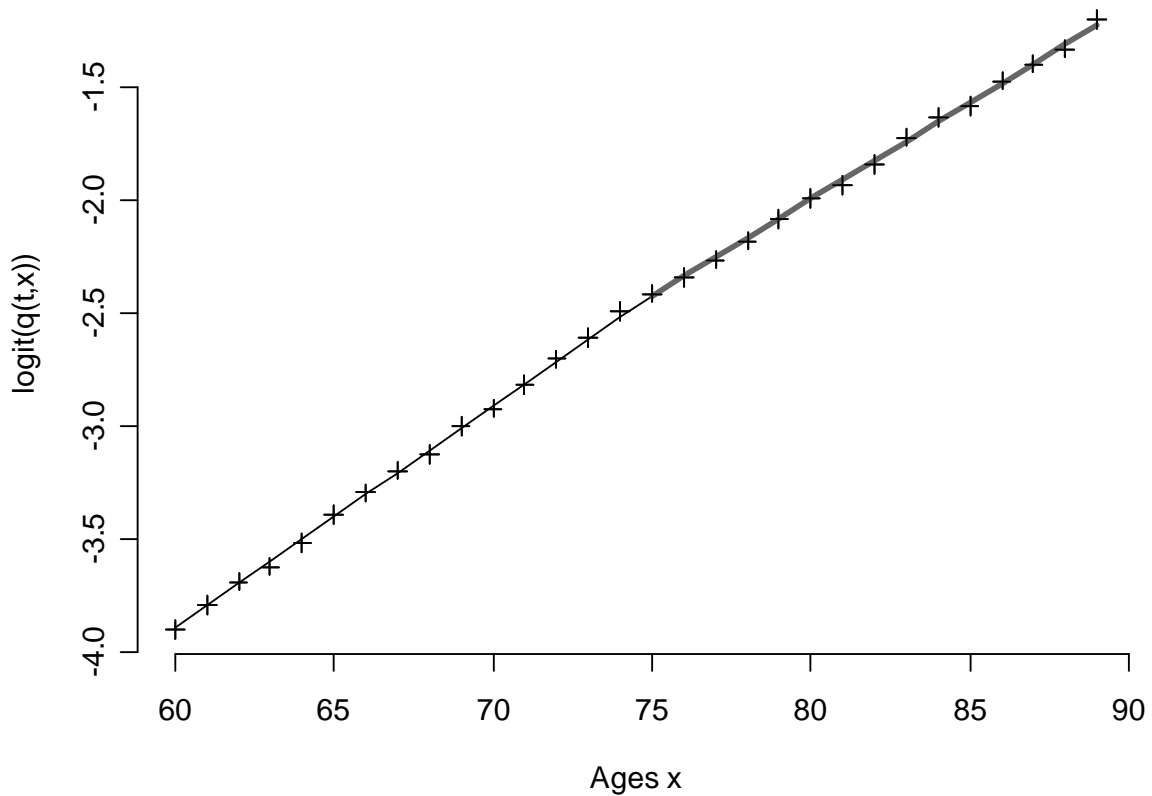


Figure 10: In year 1975

3.3.2 New Model Two (NM2)

Figure 11 displays the results obtained using two kink points, in birth years 1900 and 1920 respectively. This is NM2 (see Equation 6).

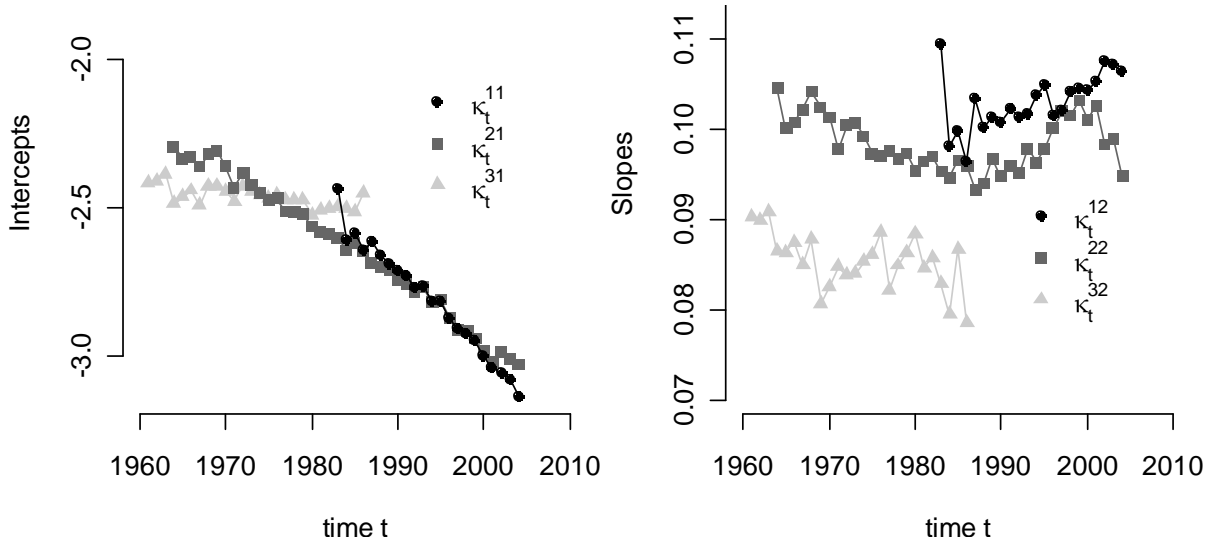


Figure 11: Parameters in NM2

The black dots and lines represent the parameters fitted for the cohorts where $t - x \geq 1920$, namely birth years in or after 1920; the grey squares and lines represent the parameters fitted for the cohorts where $1900 \leq t - x \leq 1920$, i.e. birth years are in or after year 1900, but in or before year 1920; and the light grey triangles and lines represent the parameters fitted for the cohorts where $t - x \leq 1900$, i.e. the birth years are in or before year 1900.

As in Figure 9, Figure 11 shows the intercept factors κ_t^{11} (black), κ_t^{21} (grey), κ_t^{31} (light grey) in the left picture and slope factors κ_t^{12} (black), κ_t^{22} (grey), κ_t^{32} (light grey) in the right picture. Ignoring the values with insufficient data to fit, the parameters (black dots in Figure 12) representing those who were born in or after 1920 show approximately linear trends. The outcomes imply that the gap of mortality rates of people between who were born before 1920 and those born later is becoming increasingly significant, so we need to treat each group separately.

Figure 12 gives a clear illustration about how the two kink points have impact on the fitted mortality rates. Furthermore, the improvement of goodness-of-fit can be observed as well. The logit of mortality rates in year 1985 has been taken for example in which the original linear function for fitting the data has been replaced by three linear functions (see equation NM2). From the graph, the kink points in 1900 and 1920 can be seen (ages 85 and 65 in 1985).

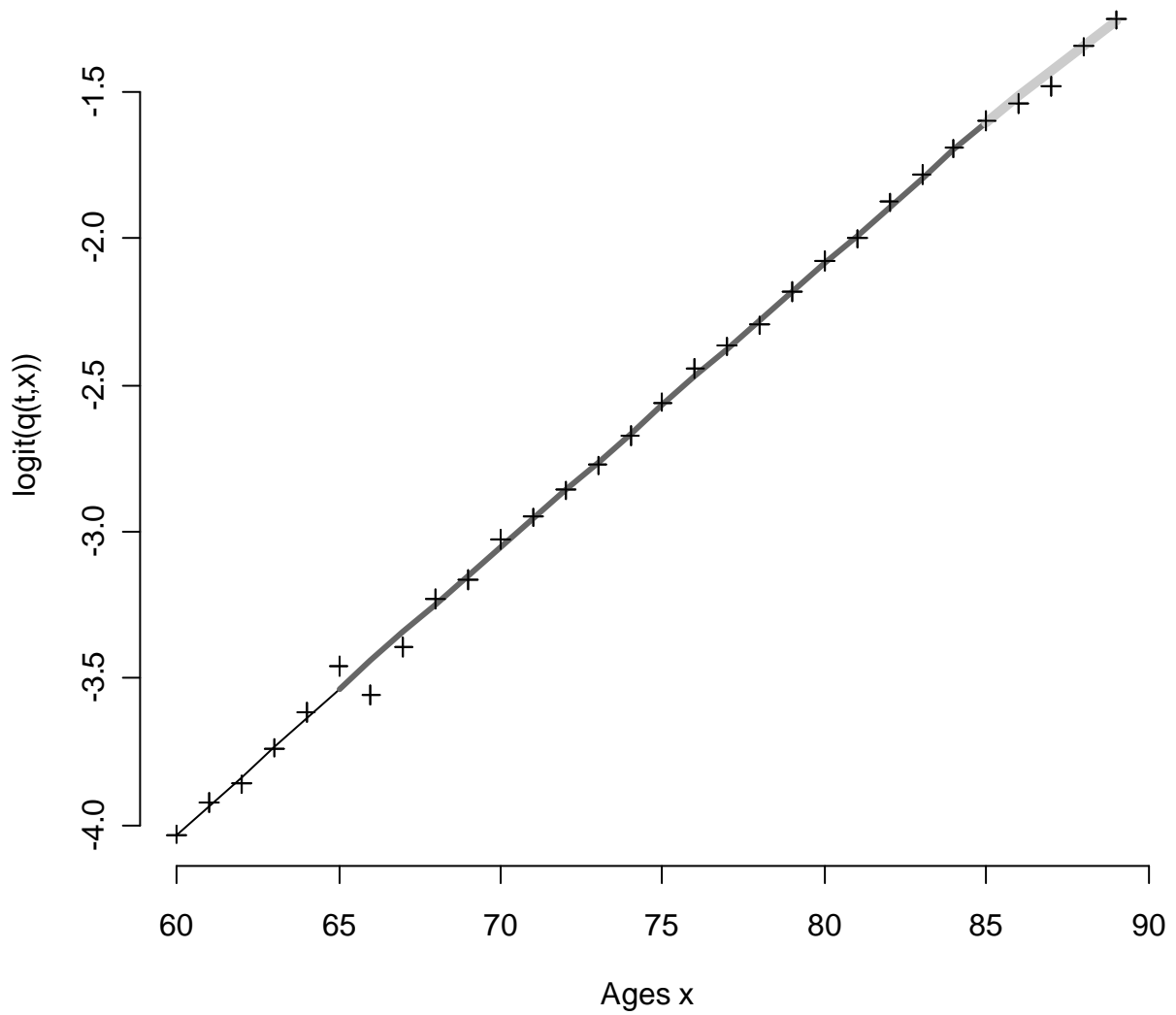


Figure 12: In year 1985

Looking more closely, it is clear that the discontinuity around 1920 is not a change in a single year; rather, there are changes spanning several years. Bearing in mind that this might be a function of data anomalies, we consider the effect of adding additional kinks to the model.

3.3.3 New Model Three (NM3)

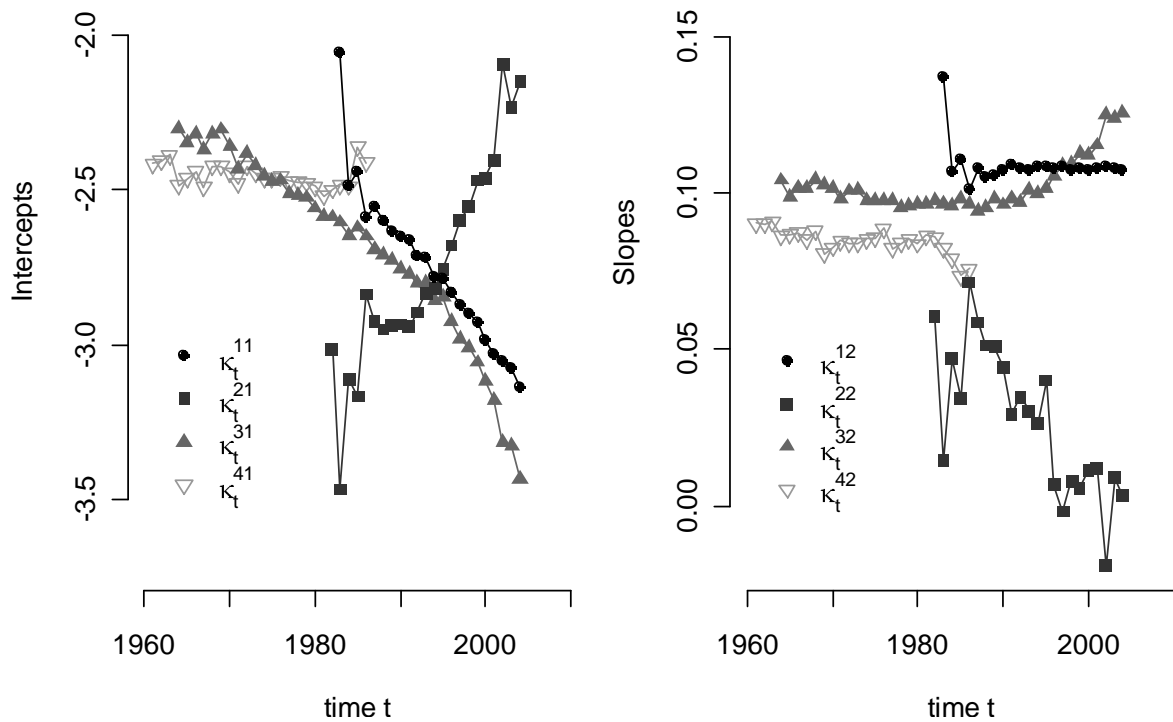


Figure 13: Parameters in NM3

Figure 13 gives the plots representing the trend of parameters in New Model Three (NM3) which allows for three kinks. The years for the three kinks are 1900, 1919, 1920. Compared with the model with two kink points, 1920 is replaced by two at 1919 and 1920 in NM3. Figure 13 illustrates clearly how the values of parameters in these two years change. The values of κ_t^{31} and κ_t^{32} show the dramatic differences. The slopes of linear functions fitting the data during the period between 1919 and 1920 tend to be negative.

Looking at the mortality rates in 1985 for example (Figure 14), it is clear that the additional kink greatly improves the fit. This picture is consistent with the direction of the cohort coefficient seen in Figure 1 and Figure 2.

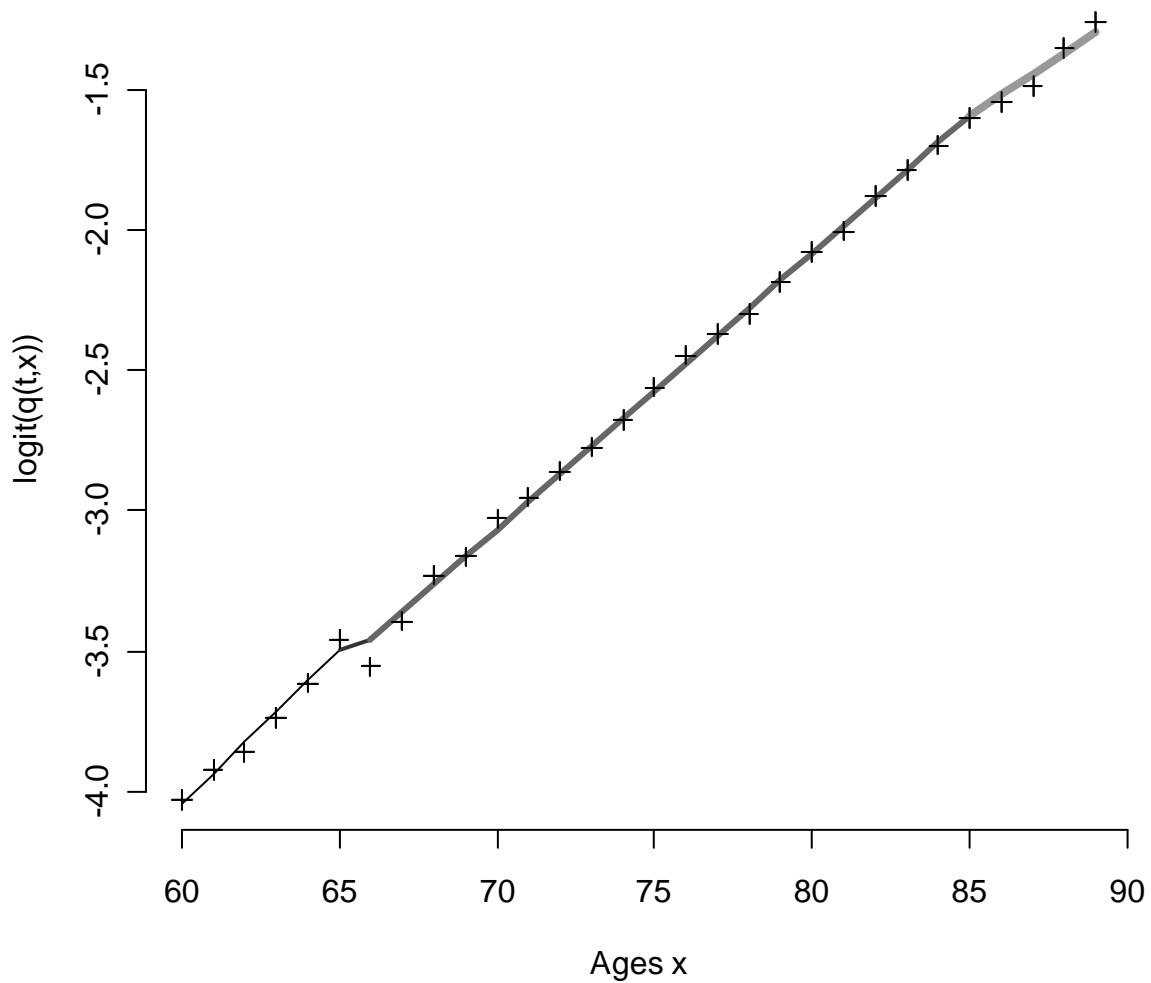


Figure 14: In year 1985

3.3.4 New Model Four (NM4)

An even closer fit can be obtained by adding a fourth kink at year 1921. In Figure 15, the parameter under NM4 and the impact of the kink points is shown.

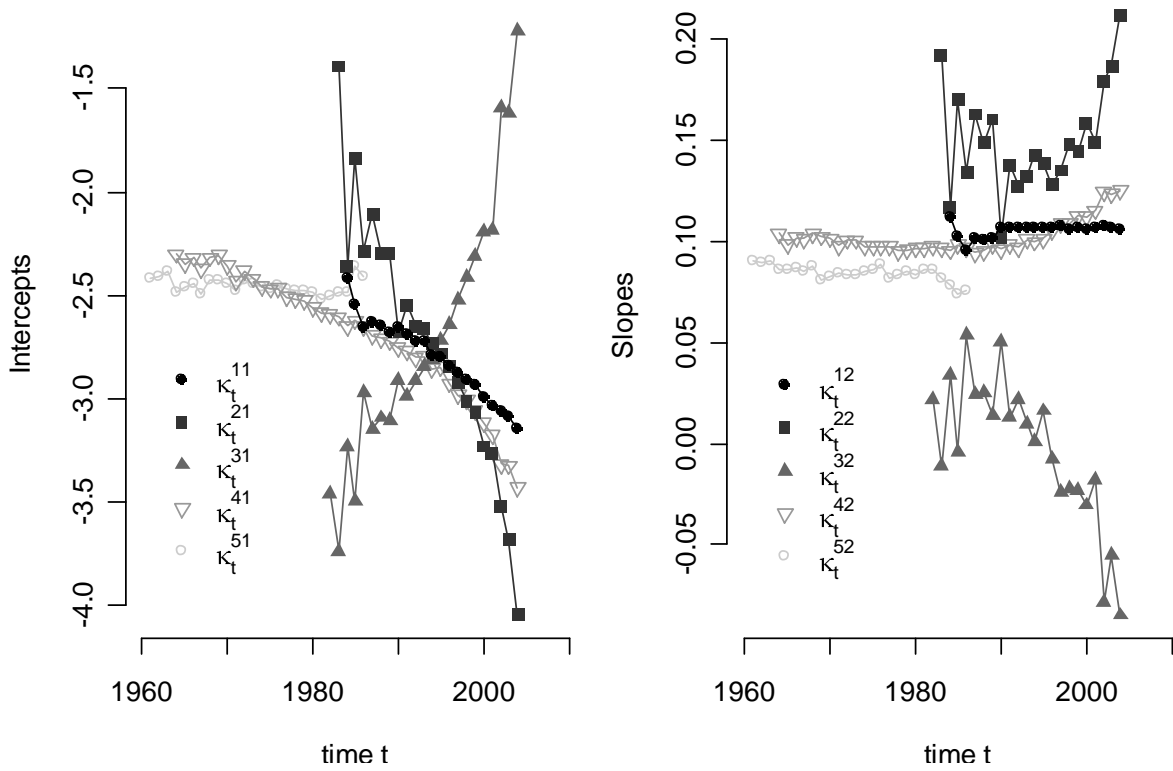


Figure 15: Parameters in NM4

Looking again at 1985, Figure 16 illustrates the effect of combining 5 different fitted linear functions into a single piecewise linear function.

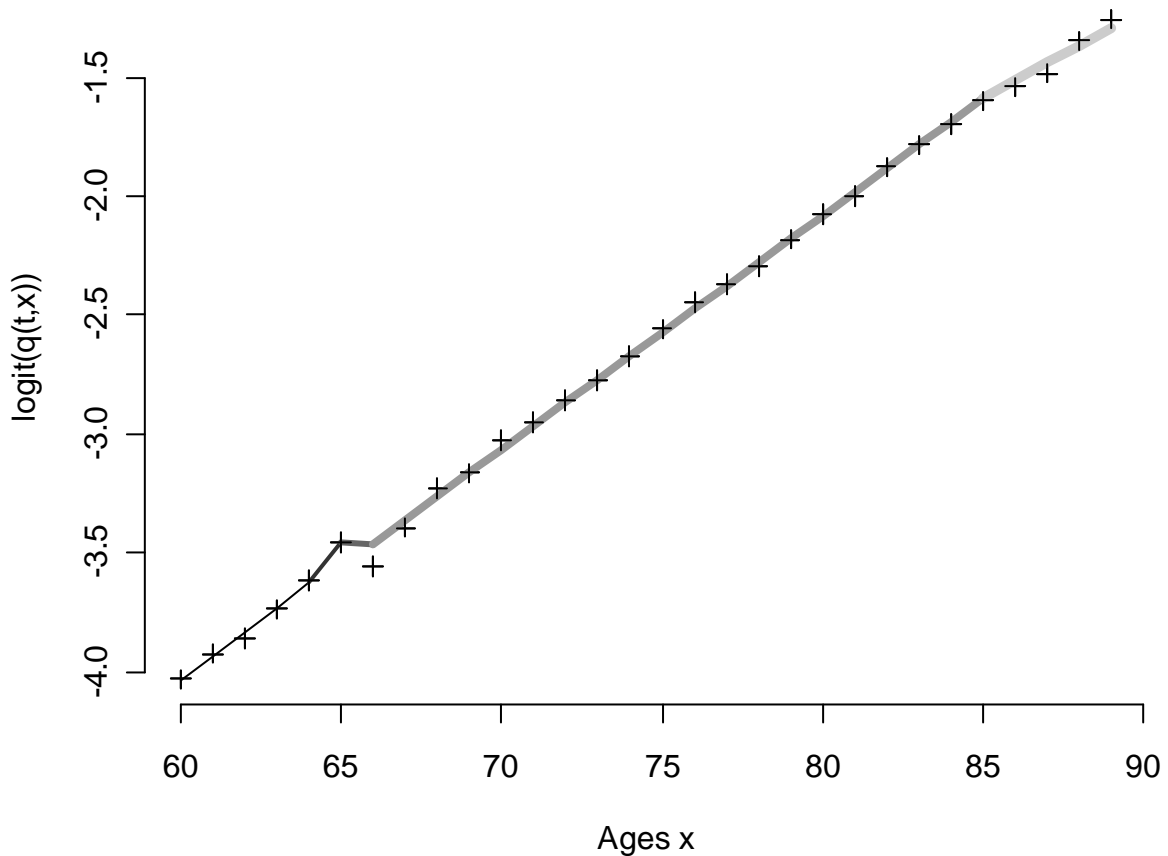


Figure 16: In year 1985

However, as before it is important to consider that this might actually be over-fitting to plot data anomalies rather than the true underlying mortality experience.

3.3.5 New Model Four (NM5)

We can add more kink points into the New Model series. Via the investigation and research above, we can find the kink points happened to match the discontinuities of cohort factor $\gamma_{t-x}^{(3)}$ in M6 (Figure 2). This also demonstrates the flexibility of the New Model family. Adding kink points into mortality models increases the extent to which rates can change from cohort to cohort. For example, consider NM5 in which the fifth kink happens in 1928. The parameters for this model are shown in Figure 17. This additional kink has an impact which is seen only in more recent years, since people born in this year did not reach 60 until 1988, so the fit of the model is shown in Figure 18 for the year 2004 rather than 1985.

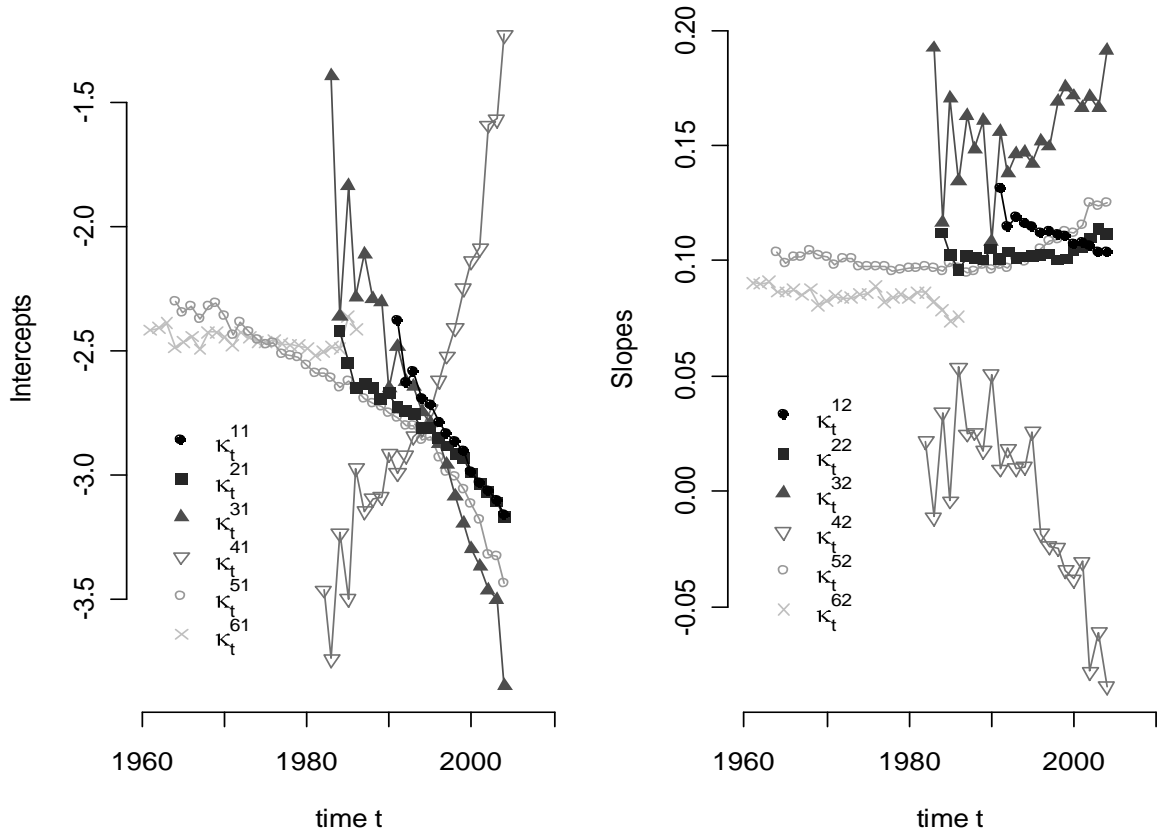


Figure 17: Parameters in NM5

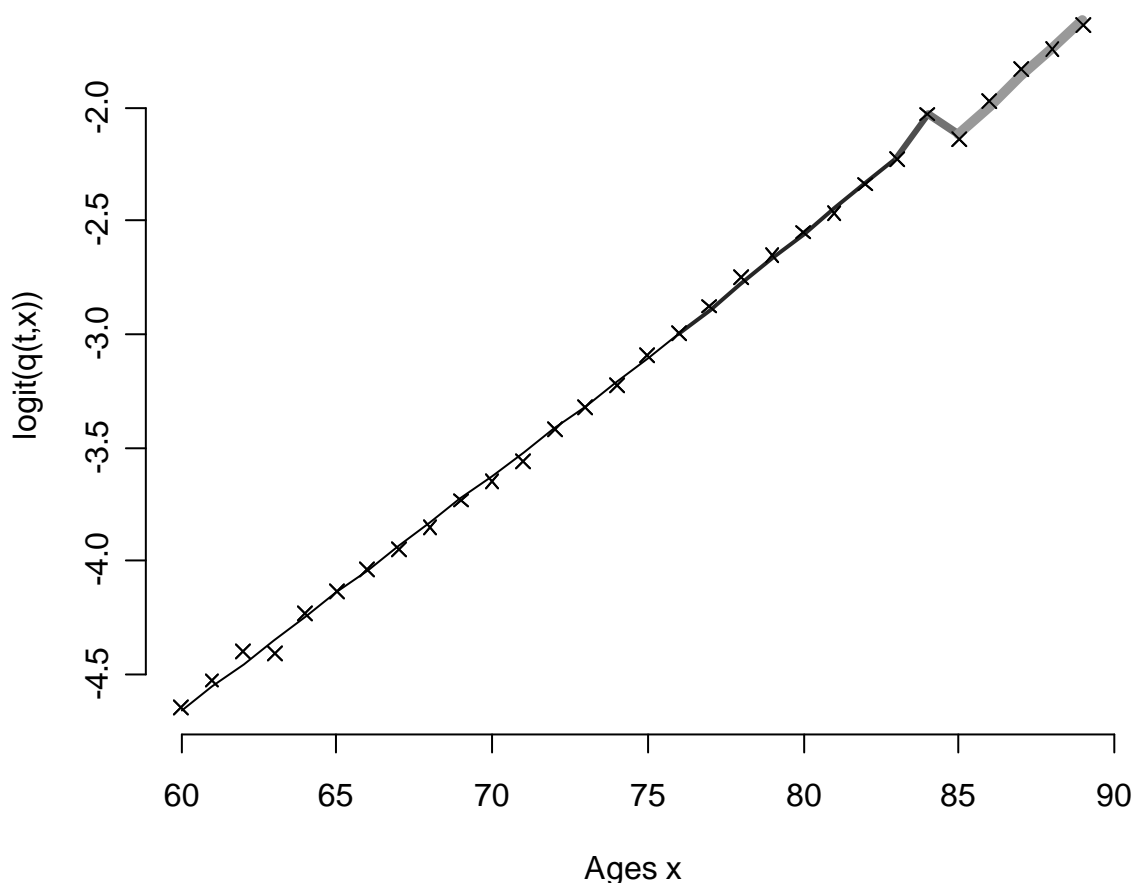


Figure 18: In year 2004

3.4 Similarities and Differences between Models

All the models in the paper are based on CBD model. However, they use not just the two factors of this model, namely slope and intercept, but also a range of kink points. The distinction between the New Models is the number of kink points in use.

Because of the existence of kink points, the new models differ from CBD model in that the latter assumes a log-linear relationship between all mortality rates in the same calendar year.

5. Analysis and Comparison

5.1 BIC & AIC

The Bayes Information Criterion (BIC) and Akaike Information Criterion (AIC) enjoy the most popularity when comparing mortality models (Sweeting 2011a). They are the criteria for model selection as a balance between goodness-of-fit and parsimony, in order to avoid over-parameterisation. Both methods have the same mechanism, that they are functions of maximum log likelihood and the number of parameters. It is obvious that fitting data better

can be fulfilled by adding more parameters in a model, but this reduces parsimony of a model and also complicates forecasting process.

The BIC is defined as: $BIC = 2l - k \log(N)$ (5.1)

while the AIC is : $AIC = 2l - 2\log(k)$ (5.2)

where l is the log likelihood, k is the effective number of parameters being estimated, and N is the number of observations.

Model s	Formula
M1	$\log(m_{tx}) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_x^{(2)}$
M2	$\log(m_{tx}) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_x^{(2)} + \beta_x^{(3)}\gamma_{t-x}^{(3)}$
M3	$\log(m_{tx}) = \beta_x^{(1)} + n_a^{-1}\kappa_x^{(2)} + n_a^{-1}\gamma_{t-x}^{(3)}$
M5	$\text{logit}(q_{tx}) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x})$
M6	$\text{logit}(q_{tx}) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_{t-x}^{(3)}$
M7	$\text{logit}(q_{tx}) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \kappa_t^{(3)}[(x - \bar{x})^2 - \hat{\sigma}_x^2] + \gamma_{t-x}^{(4)}$
M8	$\text{logit}(q_{tx}) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_{t-x}^{(3)}(x_c - \bar{x})$

Table 1

	k	l (MLE)	l (GA)	BIC	AIC	kink points
M1	102	-10664.31		-22061.54	-21332.64	
M2	203	-7819.672		-17094.88	-15643.96	
M3	144	-8468.919		-17970.34	-16942.15	
M5	88	-11228.28		-23087.54	-22460.45	
M6	159	-8000.152		-17140.35	-16004.71	
M7	202	-7792.837		-17034.04	-15590.28	
M8	161	-7937.458		-17029.31	-15879.33	
NM1	116	-9781.566	-10853.25	-22068.07	-21710.63	1900
NM2	140	-8949.035	-9667.43	-20281.62	-19344.63	1900, 1920
NM3	165	-8307.100	-8759.38	-18780.70	-17529.10	1900, 1919, 1920
NM4	188	-8109.642	-8563.73	-18704.38	-17138.25	1900, 1919, 1920, 1921
NM5	204	-7966.134	-8447.57	-18786.21	-16906.29	1900,1919,1920,1921,1928

Table 2

In Cairns et al. (2009), the comparison of seven models (Table1) has been displayed under the BIC. M4 is not included, because it is fitted using penalised splines rather than a

parametric model. The New Models use the same method in this paper. Values for the BIC and the AIC are shown in Table 2.

It is evident that new models have better values than M5 (CBD model) which does not allow for the cohort effect, but not as good as M6 which allows for a separate cohort effect for every birth year. However the new models offer a pragmatic alternative to the intricate cohort effects calculated using M6, by instead using a small number of kink points. Having said this, there is the risk that by adding additional kinks, the New Models are themselves sacrificing parsimony to fit data anomalies rather than genuine underlying factors.

M7 and M8 are far more involved than the earlier models, so of questionable parsimony, whilst according to Cairns et al (2008), the L-C model (M1) is not suitable for the demography of the UK. Cairns et al also believe that the Renshaw-Haberman model (M2) suffers from a lack of robustness. M3 is simply a special case of M2 with the coefficients of the period and cohort factors held constant.

5.2 SR

Comparing the standardized residuals (SR) of different models is another good method of model selection. The higher likelihood a model has, the lower variance of the SR it has relatively. The SR is defined as:

$$Z_{tx} = \frac{d_{tx} - e_{tx} \hat{m}_{tx}}{\sqrt{e_{tx} \hat{m}_{tx}}} \quad (5.3)$$

The variances of Standardized Residuals from M1 to M8 and the New Models are shown in Table 3. According to the outcomes, we obtain the same conclusion with the BIC and AIC test, that the New Models are a good way to deal with the cohort effect using several kink points.

	Models							
	M1	M2	M3	M4	M5	M6	M7	M8
<i>var</i>(Z_{tx})	4.1	2.2	3.7	4.3	5.9	2.4	2.1	2.3
	NM1	NM2	NM3	NM4	NM5			
<i>var</i>(Z_{tx})	4.2	4.1	2.8	2.5	2.4			

Table 3

The standardised residuals can also be shown in three dimensions, as demonstrated in Figure 19. Here, there are clear diagonal clusters of positive and negative in the first two graphs caused by the cohort effects of birth year around 1920. These lines disappear in NM3, NM4 and NM5 gradually. The two-dimensional illustrations in Figure 20 show the same illustrations in mosaic format.

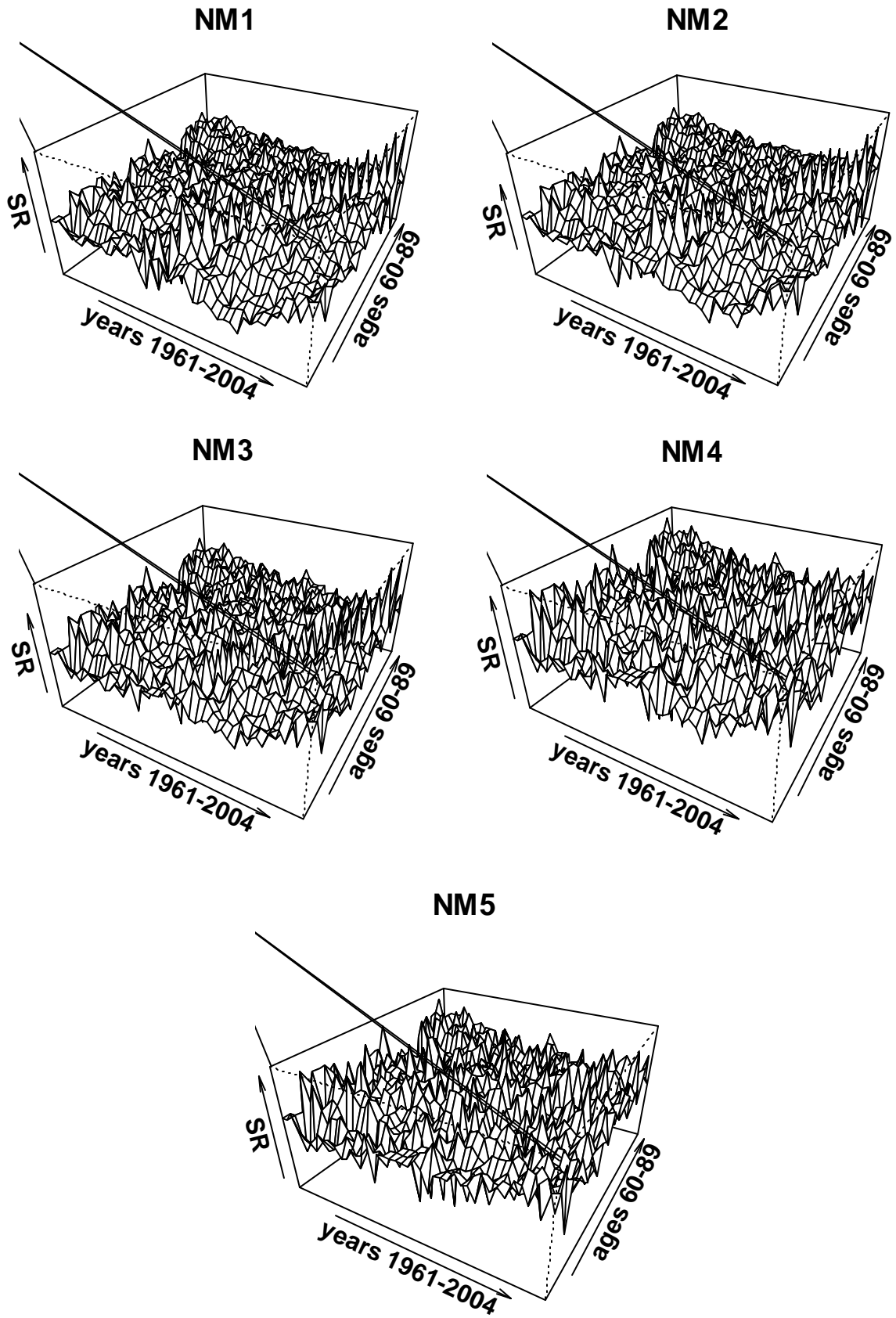


Figure 19

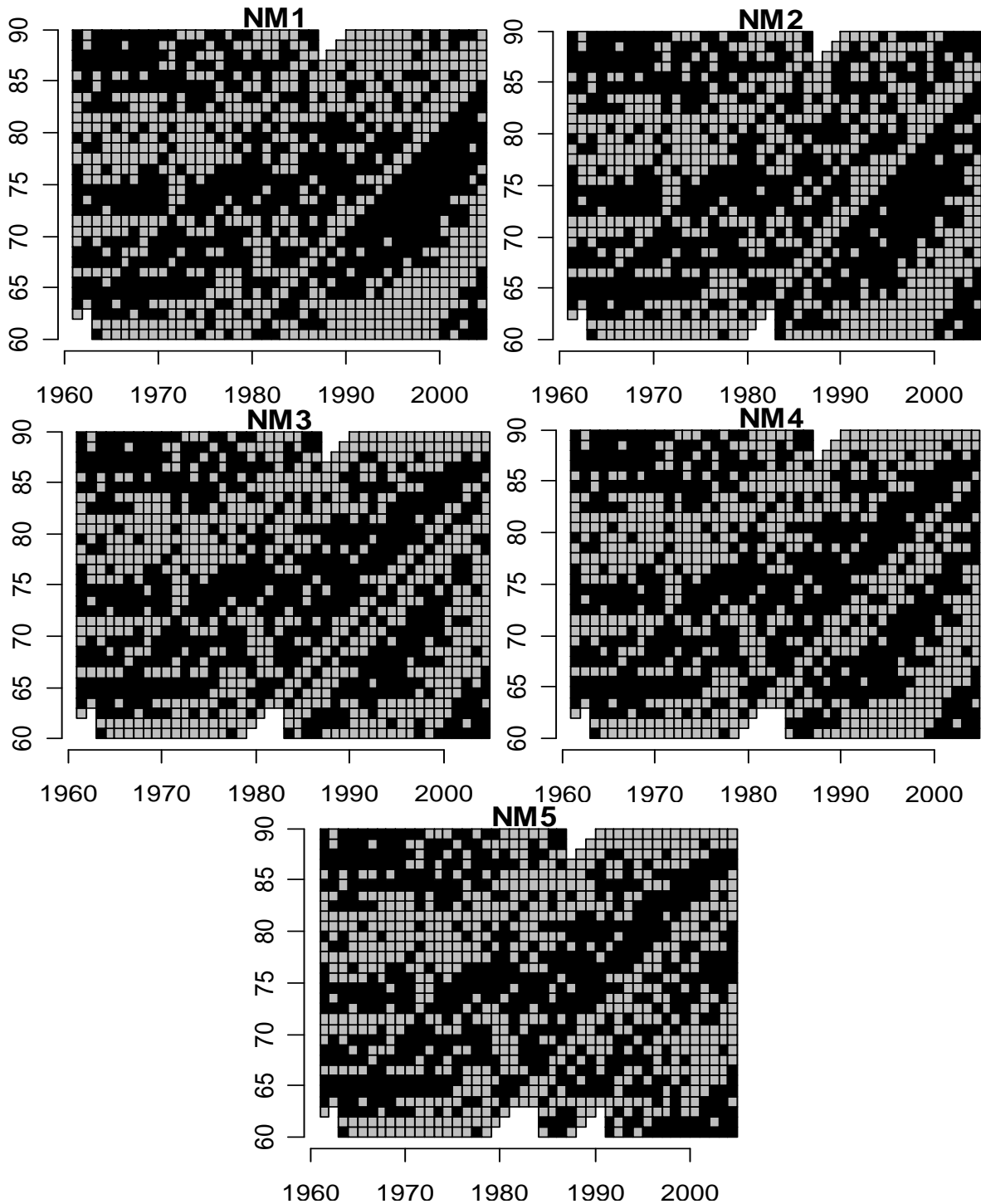


Figure 20

5.3 LRT

In the strict sense, the New Models with fewer kink points are nested within those with more kink points. For instance, CBD model is a special case of the other New Models, while NM2 implies NM1. To varying degrees, the likelihood ratio test (LRT) (Sweeting 2011a) is a good method to judge whether the nested model is acceptable or not. Take NM2 and NM1

for example, let $l1$ be the log likelihood for New Model 1 and $l2$ be the log likelihood for New Model 2. Via tests, New Model 1 possesses 116 parameters, $v1 = 116$, when New Model 2 has 140, $v2 = 140$. According to LRT:

$$2(l2 - l1) \sim \chi^2_{v2-v1} \quad (5.4)$$

The statistic $2(l2 - l1)$ follows Chi-Squared Distribution and the degree of freedom (d.f.) is $v2 - v1$. Alternatively, the p-value for the test is given as $p = 1 - \chi^2_{v2-v1}^{-1}[2(l2 - l1)]$.

Models	Nested models	$2(l2 - l1)$	d.f.	p-value
CBD model	NM1	1446.714	28	<0.000001
NM1	NM2	832.531	24	<0.000001
NM2	NM3	641.935	25	<0.000001
NM3	NM4	197.458	23	<0.000001
NM4	NM5	143.508	16	<0.000001

Table 4

This suggests that each kink point added significantly improves the fit of the model. However, from the point of view of projections, it is at least as important to consider the parsimony of the model.

6. Conclusions

In the paper, obtaining the inspiration from the smoothness of the CBD model, New Models replace the smoothness across ages in same year with one or more linear functions in each year to fit the data, enhancing the goodness-of-fit which can be seen from the values of SR especially. From the perspective of parsimony, the new models do create more parameters than the original M5 model; however, the methods shown here better reflect the way in which mortality rates develop over time.

It is, though, important not to over-parameterise these models. Some apparent kinks, particularly those occurring around 1920, are likely to arise as a result of data anomalies. Even though using additional parameters to model these characteristics might give a significantly better fit, a model with fewer parameters such as NM3 or even NM2 might give a realistic basis for projections.

References

- Bajpai, P. & Kuma, M. (2008) Genetic Algorithm – an Approach to Solve Global Optimization Problems. *Indian Journal of Computer Science and Engineering*, Vol 1 No. 3 199-206
- Booth, H. & Tickle, L. (2008). Mortality Modelling and Forecasting: a Review of Methods. *Annals of Actuarial Science*, 3, 3-43

- Brouhns, N., Denuit, M. & Vermunt, J. K. (2002). A Poisson Log-Bilinear Regression Approach to the Construction of Projected Life Tables. *Insurance: Mathematics and Economics*, 31: 373–93.
- Cairns, A. J. G., Blake, D. & Dowd, K. (2006b). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, 73: 687–718.
- Cairns, A. J. G., Blake, D. & Dowd, K., Coughlan, G. D., Epstein, D. & Khalaf-Allah Marwa (2008). Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models. *Pi Pension Institute Discussion Paper*, PI-0801.
- Cairns, A.J.G., Blake, D., Dowd, K. (2008). Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal* 108(2-3): 79-113.
- Cairns, A. J. G., 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. *Pi Pension Institute Discussion Paper*, PI-0701.
- Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28, 591-605.
- Continuous Mortality Investigation (CMI) (2005). Projecting Future Mortality: Towards a Proposal for a Stochastic Methodology. *Working paper* 15.
- Continuous Mortality Investigation (CMI) (2007). Stochastic projection methodologies: Lee-Carter model features, example results and implications", *Working paper* 25.
- Herman J. B. (March 12, 2006). Information Criteria and Model Selection. *Pennsylvania State University March*, 12, 2006
- Koissi, M. C., Shapiro, A. F. & Högnäs, G. (2006). Evaluation and Extending the Lee-Carter Model for Mortality Forecasting: Bootstrap Confidence Intervals. *Insurance: Mathematics and Economics*, 38: 1-20.
- Lee, R. D. & Carter, L. R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87: 659–75.
- Renshaw, A. E. & Haberman, S. (2003). Lee-Carter Mortality Forecasting with Age-Specific Enhancement. *Insurance: Mathematics and Economics*, 33: 255–72.
- Renshaw, A. E. & Haberman, S. (2006). A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors, *Insurance: Mathematics and Economics*, 38:556-570.
- Sweeting, P. J. (2011). A Trend-Change Extension of the Cairns-Blake-Dowd Model. *Annals of Actuarial Science*, page 1 of 20 [doi:10.1017/S1748499511000017](https://doi.org/10.1017/S1748499511000017).

Sweeting, P. J. (2011a). *Financial Enterprise Risk Management*. Cambridge University Press, 1 Sep 2011

Willets Richard (2003). *The Cohort Effect Insights and Explanations*. *Willets Consulting Limited* 2003.