Modelling heterogeneity in human populations: Towards a better understanding of overall mortality patterns for life insurance business

Thesis Proposal for HDip Data Science and Analytics

Paul Christopher, M.A.

Thesis Supervisor

Damian Conway

Date

2 May, 2022

Abstract

This paper considers the various methods used to model mortality patterns amongst a population, focusing on a recent heterogeneous Gompertz model that has been proposed by Avraam et al and applies it to Irish data. Parameters will be estimated and tested for this model from data obtained from various sources within and outside Ireland. Candidate models are constructed and their robustness assessed.

Contents

List of Figures

List of Tables

1 Introduction

Life insurance depends heavily on assessing how the demographics of a population behave and evolve. Of particular importance are mortality statistics. As Atkinson and Dickson [1] put it: "the raw material of demography is the collection of data from actual populations. These data are used to construct models describing the survival experience of populations of a certain type and these models are then used to estimate the experience of other comparable populations, or individuals within them."

A typical question of general interest might be: how many policyholders will die in the next year? This random variable is commonly referred to as $T(x)$, shorthand for the length of the future lifetime of an individual now aged exactly x years of age. Accordingly, $T(0)$ represents the total lifetime of an individual within a population under study. Another interesting question from a life insurer's point of view may be: what is the probability that a randomly chosen person will live to 100?

Actuaries have developed a tool, known as a life table, to answer such questions. It is constructed from population mortality to facilitate the calculation of probabilities of death or survival of an individual drawn from that population. Typical elements of the life table are l_x, d_x and q_x and they are defined and explained in more detail *infra*.

As Atkinson and Dickson [1] states: "in order to construct a mathematical model which is representative of human survival experience, it is necessary first to examine the actual experience of a group of individuals, from which to derive the model." This data is typically obtained from census data, registration (of births and deaths) records and proprietary data, typically from insurance companies.

Life tables may also contain values for other random variables, such as the survival function, $s(x)$, which represents the probability that a newborn will survive for at least x number of years.

Another random variable, the one that this paper is primarily concerned with, is used to describe the rate of death at any moment: the force of mortality at age x, denoted by μ_x. It is defined as

$$\mu_x = \lim_{\Delta x \to 0} \frac{1}{\Delta x} Pr(x < T(0) \leq x + \Delta x | T(0) > x) \tag{1}$$

$$= \frac{-s'(x)}{s(x)} \tag{2}$$

Mortality rates exhibit a fairly typical pattern, which show high initial mortality rates around birth, representing infant mortality, falling to a low around early teenage years, showing a 'hump' in the early 20s, mainly due to accidental and violent deaths and then exhibiting an increase upto about 80 or 90 years, where the rate slows again.

This pattern appears to be consistent across time and cohorts.

In the html version of this document, the animation at Figure 1 shows the evolution of mortality curves in Ireland from 1950 to 2017.[1] The colors of the curves follow the order of a rainbow, with the oldest data in red and the most recent data in violet.[2]

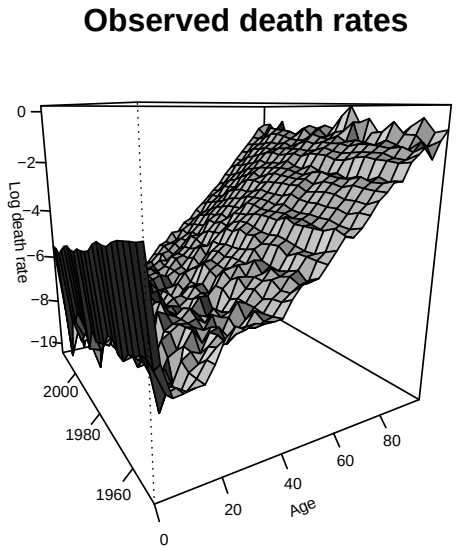


Figure 1: Mortality rates: Ireland: 1950 to 2017

A further random variable contained in life tables, that is constructed from mortality rates, is the expectation of life at age x, denoted e_x. One of the most common values derived from this is the expectation of life at birth, e_0. Figure 2 shows how the expectation of life at birth has increased over the years in Ireland and how it varies, according to gender.

To enable analytical representations of mortality rates, a mathematical model of mortality is required. There are two long-established mathematical models which have

[1] In the pdf version, a static 3D plot of the same thing is shown.

[2] Data was pulled directly from Human Mortality Database (HMD) using the `demography` package into R and the animation was created using the `animation` package.

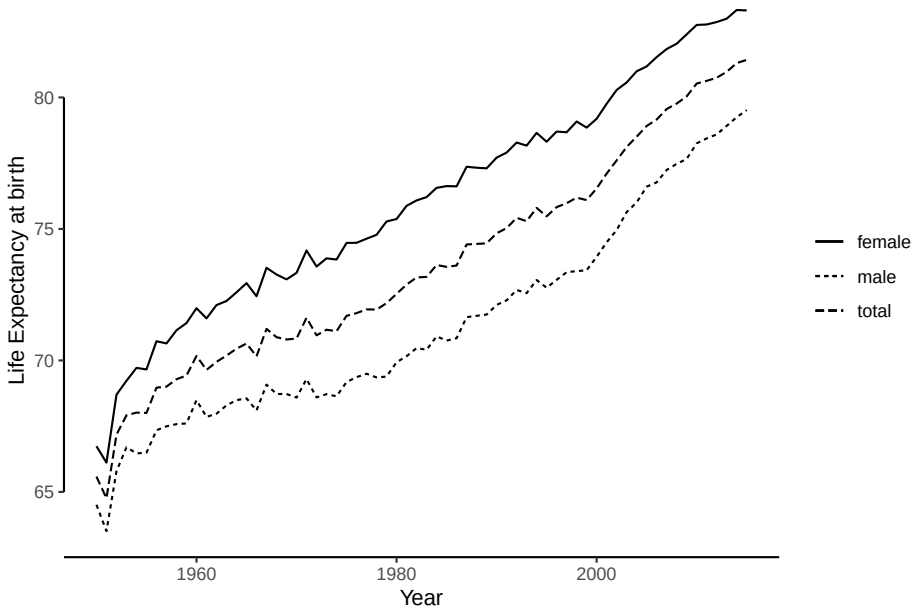


Figure 2: Expectation of life at birth in Ireland for male, female and total population: change between 1950 - 2015. *Source*: Human Mortality Database.

been suggested as being appropriate for use over limited age ranges:

$$\text{Gompertz law: } \mu_x = \alpha e^{\beta x} \tag{3}$$

and

$$\text{Makeham law: } \mu_x = \alpha e^{\beta x} + \gamma \tag{4}$$

where μ_x is the mortality rate at age x and α, β and γ are suitably chosen parameters.

However, as Atkinson and Dickson [1] states, "the curve of the mortality rate (or force of mortality, μ_x at age x) for humans cannot be adequately modelled by a single mathematical function, because it shows different characteristics over different age ranges."

A recent development has been to treat the overall population as being composed of heterogeneous groups. These different groups are assumed to be internally homogeneous, that is the individuals comprising it are regarded as sharing the same characteristics and parameters of a Gompertz-type model. The model for the overall population is then constructed by additively amalgamating Gompertz-type models, as specified above, for each such sub-group into one overall mathematical model. An example of where this has been done is in Avraam, Magalhaes, and Vasiev [2].

The Gompertz model for each sub-group will be specified by two parameters: $\alpha, \text{and}, \beta$. The overall model would then be specified by $3 * n$ parameters, where n is the number of sub-groups and another parameter, the weighting for each sub-group, ρ_j, is used to give an overall model as follows:

$$\mu_x = \sum_{j=1}^{n} \rho_{j,x} \alpha_j e^{\beta_j x} = \sum_{j=1}^{n} \rho_{j,x} m_{j,0} e^{\beta j x} \quad \text{where } \sum_{j=1}^{n} \rho_j = 1 \tag{5}$$

where the sub-index j indicates the j-th out of n subpopulations, $m_{j,0}$ is the central death rate at age 0 of sub-population j, and β_j is the mortality coefficient which gives the rate of change of mortality with age. The weights $\rho_{j,x}$ are fractions formed by each sub-population j at age x in the entire population and their sum is equal to unity at all ages. The mortality rate at age 0 of the sub-population j is equal to α_j and thus the relationship $m_{j,0} = \alpha_j$ leads to the last term in the equation.

This paper will be primarily concerned with estimating and testing these parameters from data obtained from various sources within and outside Ireland and constructing suitable candidate models and analysing their robustness, *inter alia*, over time.

2 Demographic background and definitions

2.1 Life tables

Life tables are one of the standard tools in a demographer's or actuary's armoury. They provide useful information for mortality forecasting and generate simple summary statistics that are useful for comparisons on such metrics as life expectancy.

The basic life table is single decrement, whereby all forms of death are lumped together; complete, in that it covers all age classes and; is of the cohort type, meaning that it provides a longitudinal perspective from the moment of birth through consecutive ages until the death of all individuals in the original cohort.

A single decrement process is a process that can be dichotomised with respect to its outcomes. For example, in the context of mortality a person alive in period x can either move into the next period, x+1, alive or they die and do not. It can be visualised as in Figure 3.

An extract from the Irish CSO female period Life Table for the years 2015 - 2017 is shown at Figure 4 for illustrative purposes. Only the ages from 0 to 8 and then from ages 97 to 105 are shown. The full life table would have the ages between those ranges included as well.

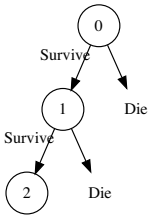


Figure 3: Single Decrement Process

Table 2 Irish Life Table No. 17[2], female period life expectancy by age, 2015 - 2017

Age x[1]	l_x[1]	d_x[1]	p_x[1]	q_x[1]	L_x[1]	T_x[1]	e^0x[1]	Age x
Exact age of person	Number of persons surviving	Number of deaths	Probability of surviving a year	Rate of mortality	Population expected	Expected number of person years lived	Life expectancy at age x	
0	100,000	304	0.9969598	0.00304020	99,848	8,343,780	83.44	0
1	99,696	22	0.9997842	0.00021585	99,685	8,243,932	82.69	1
2	99,674	7	0.9999332	0.00006683	99,671	8,144,247	81.71	2
3	99,668	5	0.9999476	0.00005239	99,665	8,044,575	80.71	3
4	99,663	6	0.9999357	0.00006425	99,659	7,944,910	79.72	4
5	99,656	8	0.9999239	0.00007613	99,652	7,845,251	78.72	5
6	99,649	6	0.9999355	0.00006452	99,645	7,745,599	77.73	6
7	99,642	5	0.9999470	0.00005301	99,640	7,645,953	76.73	7
8	99,637	5	0.9999536	0.00004637	99,635	7,546,314	75.74	8
97	6,484	1826	0.7184240	0.28157596	5,571	17,082	2.63	97
98	4,658	1400	0.6993552	0.30064477	3,958	11,511	2.47	98
99	3,258	1037	0.6816024	0.31839759	2,739	7,553	2.32	99
100	2,221	743	0.6656126	0.33438743	1,849	4,814	2.17	100
101	1,478	515	0.6518133	0.34818666	1,221	2,964	2.01	101
102	963	346	0.6405984	0.35940165	790	1,744	1.81	102
103	617	227	0.6323156	0.36768442	504	953	1.54	103
104	390	136	0.6524701	0.34752987	322	450	1.15	104
105	255	255	0.7895501	0.21044991	127	127	0.50	105

[1]See below and background notes.

x the exact age of the person, that is on his or her birthday.

l_x the number of persons surviving to exact age x out of the original 100,000 aged 0.

d_x the number of deaths in the year of age x to x+1 out of l_x persons who enter that year.

p_x the probability of surviving a year, or the ratio of the number completing the year of age x to x+1 to the number entering on the year.

q_x the rate of mortality, the probability of dying in a year, or the ratio of the number of deaths in the year of age x to x+1 to the number entering on the year.

L_x the population to be expected according to the Life Table aged between x and x+1 years, assuming deaths occur evenly over the year.

T_x the expected number of person years to be lived by the survivors at age x.

e^0x life expectancy at age x for each person surviving, or the total future life time in years which will on average be passed through by persons aged exactly x.

Figure 4: Extract from CSO female period Life Table for 2015-2017.

A life table has several functions or parameters set out in columns for each age from 0 to ω,[3] each age being a row of the table. The primary functions are as follows:

2.1.0.1 Survival l_x The most basic life table function is defined as the expected number of individuals alive at age x in a specific population or cohort from a limited population of l_0 individuals aged 0, or the surviving members of the cohort. l_0 is termed the 'radix' and is usually, arbitrarily, set at 100,000. There are two approaches to calculating this function. A *cohort* life table is the sequence $l_0, l_1, ..., l_\omega$ provided by statistical evidence, *i.e.* a longitudinal observation of the actual numbers of individuals alive at age 1,2,. . . ,ω, out of a given initial cohort, l_0. The integer age, ω is the limiting age, that is, the age such that $l_\omega > 0$ and $l_{\omega+1} = 0$.

Conversely, the *period* life table is obtained by assuming that the statistical evidence consists of the frequency of death at various ages, observed through a given period, for example, one year. Assume that the frequency of death at age x is an estimate of the probability q_x, then, for $x = 0, 1, ..., \omega - 1$ define: $l_{x+1} = l_x(1 - q_x)$ with l_0 assigned, for example, 100,000. Hence, l_x is the expected number of survivors out of a notional cohort initially consisting of l_0 individuals.[4] An important assumption underlying the recursive formula for period life tables is that the mortality pattern does not change from a given period (say, one year) where the q_x's were estimated. This is not a realistic assumption, given the general decline in mortality over the last century. For that reason, cohort life tables are generally preferred.

2.1.0.2 Age-specific survival, p_x The second life table parameter, denoted p_x, is age-specific (or period) survival probability defined as the probability of surviving from age x to $x + 1$. It is computed from the l_x schedule as: $p_x = \frac{l_{x+1}}{l_x}$. If T_x denotes the random variable representing an individual's remaining lifetime at age x, the probability that the individual will surivive an additional h years is: ${}_hp_x = \mathbb{P}[T_x > h]$.

2.1.0.3 Age-specific mortality, q_x The complement of the period survival schedule is the age-specific (period) mortality schedule denoted q_x and defined as the probability of an individual dying in the interval x to $x + 1$:

$$q_x = \frac{l_x - l_{x+1}}{l_x} = 1 - \frac{l_{x+1}}{l_x} = 1 - p_x \tag{6}$$

[3]Where ω is the limiting age of the mortality table, 105 in the case of the CSO Life Table for the years 2015-2017.

[4]Pitacco [15].

The sum of p_x and q_x must equal unity.

2.1.0.4 Age-specific deaths, d_x is the fraction of the birth cohort that dies in the interval x to $x+1$. It is computed as:

$$d_x = l_x - l_{x+1} \ . \tag{7}$$

This parameter represents the incremental decrease in l_x due to mortality within the age interval. It is evident that $\sum_{x=0}^{\omega} d_x = l_0$.

2.1.0.5 Expectation of life, e_x Age-specific expectation of life, denoted e_x, is defined as the umber of years remaining for the average individual age x. The value of e_x is computed as the sum of life years remaining to a cohort age x normalised by the fraction alive at age x. This computation can be demonstrated using the number of life years remaining to a newborn (*i.e.* e_0) determined using the l_x schedule as:

$$e_0 = \frac{l_0 + l_1}{2} + \frac{l_1 + l_2}{2} + ... + \frac{l_{\omega-1} + l_\omega}{2} \tag{8}$$

$$= \frac{1}{2}(l_0 + 2l_1 + 2l_2 + ... + l_\omega) \tag{9}$$

$$= \frac{1}{2} + \sum_{x=1}^{\omega} l_x \quad \text{(given that } l_0 \text{ always equals unity.)} \tag{10}$$

2.1.0.6 Central death rate, m_x For a given population or cohort, the central death rate at age x during a given period of 12 months is found by dividing the number of people who died during this period while aged x (that is, after they had reached the exact age x but before reached the exact age $x+1$) by the average number who were living in that age group during the period. The simplest case is for a stationary population, in which the number of people who leave the age group during the year (either by reaching age $x+1$ or by dying) is exactly balanced by the number who enter the age group on reaching their x-th birthday. In this simple case, the number living in the age group is constant throughout the year, say n_x. If the number of deaths at age x is d_x, then the central death rate is given by: $m_x = \frac{d_x}{n_x}$.

For an observed population which is not stationary, d_x can still be measured by obtaining the average number living at age x during the year from a cohort with a known survival function l_x as follows, where the average number living at age x during

the year is:

$$L_x = \int_0^1 l_{x+t}\, dt \tag{11}$$

$$\tag{12}$$

and the central death rate is given by:

$$m_x = \frac{d_x}{L_x} \tag{13}$$

According to the methodological notes from mortality.org: "For period life tables, the central death rate m_x is used to compute probabilities of death q_x."

The central death rate is closely related to the age-specific mortality rate, q_x, which can be approximated as: $q_x \simeq 1 - e^{-m_x}$. The values of q_x and m_x are similiar at most ages. However, m_x usually exceeds q_x at very young and very old ages. This is because the within-age distribution of deaths at these extremes tend to be skewed left (*e.g.* most infant deaths occur shortly after birth) thus reducing the within-interval life-years lived by individuals who die (*i.e.* the average is not at the age 0 to 1 interval mid-point).

2.1.1 Life Table Metrics

Apart from these basic parameters of the life table, there are some additional metrics associated with the life table. They are as follows:

2.1.1.1 Life table aging rate, k_x The Life Table Aging Rate (LAR) parameter, denoted k_x, is defined as the rate of change in age-specific mortality with age. This measure is based on relative rather than absolute rate of change in mortality with age.

$$k_x = \ln(m_{x+1}) - \ln(m_x) \tag{14}$$

where m_x denotes the central death rate. The life table aging rate is a measure of the slope of mortality with respect to age.

2.1.1.2 Life table entropy Life table entropy provides a metric to evaluate improvements in mortality and survival in a population. If all individuals die at exactly the same age the l_x schedule is rectangular whereas if all individuals have exactly the same probability of dying at each age the l_x schedule decreases geometrically. The distribution of deaths by age varies greatly between these two patterns. A measure of

this heterogeneity known as entropy, H, can be computed as:

$$H = \frac{\sum_{x=0}^{\omega} e_x d_x}{e_0} \tag{15}$$

The sum of products $e_x d_x$ in the numerator can be viewed as either:

- the weighted average of life expectancies at age x;
- the average days of future life that are lost by the observed death;
- the average number of additional days an individual could expect to live, given a second chance on life.

The denominator is the expectation of life at birth, e_0, which converts an absolute effect into a relative effect.

Entropy can be interpreted as:

- the proportional increase in life expectancy at birth if every individual's first death were averted;
- percentage change in life expectancy produced by a reduction of 1% in teh force of mortality at all ages;
- the number of days lost due to death per number of days lived.

3 Modelling generally

At a very general level, mathematical models can be categorised according to their intended purposes, most generally between predictive models and explanatory models. Often, the purposes overlap. One can say, at a general level, that good explanatory models require relatively complex models, while good prediction may be better served by simpler models. One can also generally say that a model that is too simple will distort the message of the data, resulting in bias; while one that is too complex will be hard to estimate well, leading to inflated uncertainties. It is often said that "it is not helpful to ask whether a model is true; rather, one should ask whether it is a good description."[5] Howard Emmons put it another way when he stated that the challenge in modelling is "not to produce the most comprehensive descriptive model but to produce the simplest possible model that incorporates the major features of the phenomenon of interest".

This paper will focus on a recently developed technique for modelling heterogeneithy in human populations developed by Avraam, Magalhaes, and Vasiev [2], which illustrates the above adage and examine its applicability to human populations in Ireland and elsewhere.

[5] Christie [6].

A further dichotomy that exists in relation to the construction and use of mathematical models is that between projection and forecasting or, to put in another way, projection as distinct from prediction. Projection answers a variety of forms of questions such as: what will be the ultimate age distribution if present death rates continue. Forecasting, on the other hand, is said to involve statements intended to apply to a real rather than to a hypothetical future (Keyfitz and Caswell [12]).

In the context of models of human mortality, a second general dichotomy can be between discrete models and continuous models.

Mathematical modelling, in particular with respect to models of human mortality, involves two steps:

- specification of the model, and;
- estimation of the parameters of that model.

This will generally involve comparing competing models and their parameters. The data used to perform these comparisons, in the context of the within paper, will generally come from mortality statistics on human populations collected by various agencies, usually state, but also commercial, such as life insurance companies in the form of life tables.

3.1 Data sources

Elandt-Johnson and Johnson [7] state that: "A major subdivision of mortality data is between data relating to populations under more or less uncontrolled conditions (such as statistics of human deaths in a state) and those observed under controlled conditions of a more or less experimental nature (as in a clinical trial)." To some extent the mortality data held by insurance companies is more properly categorised as the latter rather than the former type.

The three primary data sources used in this papers are:

- from the Irish Central Statistics Office (CSO);
- from the Human Mortality Database (HMD) (mortality.org), and;
- from the Irish Life insurance company.

The CSO produces and publishes various life tables from Ireland with data from between 1871 to 2016, usually at ten year intervals.[6] Most of the life tables are based on five year age intervals, as opposed to one year age intervals, particularly for the older series. This is a shortcoming of much of the CSO data. However, its more recent life tables provide data at one year age intervals.

[6] data.cso.ie.

The Human Mortality Database (HMD) was created to provide detailed mortality and population data to researchers, students, journalists, policy analysts, and others interested in the history of human longevity. The project began as an outgrowth of earlier projects in the Department of Demography at the University of California, Berkeley, USA, and at the Max Planck Institute for Demographic Research in Rostock, Germany.[7] It provides open, international access to these data. At present the database contains detailed population and mortality data for 41 countries or areas, including Ireland. Currently, its period data generally spans the years between 1950 and 2017 and its cohort data generally spans the years 1871 to 1987. It provides data in various combinations of age interval by year intervals from 1x1 to 5x10. There are many convenient packages in R to download and manipulate this data directly into R.[8]

In addition to the aforementioned open access data sources, the author was provided with proprietary data from the Irish Life insurance company comprising life tables from the years 1935 to 2015, at ten year intervals. These life tables are based on observations from this company's policyholders.

4 Modelling Mortality

Construction of mathematical models of mortality enable analytical representations of mortality rates.[9] Mortality modelling and the concepts of actuarial aging, relative risk, odds ratio, sex mortality differentials, cause-specific and all-cause mortality and average lifetime mortality are all based on risk concepts. Mortality is the single most important concept in actuarial analysis as it notes and quantifies the risk of dying. Although life table parameters l_x, q_x, p_x, d_x and e_x are all derived from the same data, the risk of dying, q_x, underlies the actuarial outcome of all cohorts.

The importance of mortality as a metric is due to a number of factors, not least, the fact that death is an event which allows it to be modelled as a single decrement process; one can disaggregate death by cause; it allows for demographic independence and; one can create mathematical models, such as the Gompertz model, relatively easily to describe and predict the process.

4.1 Discrete and continuous mortality metrics

It is important to make a distinction between discrete and continuous mortality metrics.

[7] mortality.org.
[8] Examples include the `demography`, `HMDHFDplus` and `MortalityLaws` package.
[9] Atkinson and Dickson [1]

4.2 Discrete Mortality

There are two categories of discrete mortality metrics:

- **Probability** - risk relative to the total number exposed (*e.g.* a 0.1 probability of dying)
- **Rate** - deaths relative to life-years lived in a specific period (*e.g.* 1 death *per* 10 life-years lived). The difference between them is evident in their algebraic definitions:

$$q_x = \frac{\text{no. dying in interval x to x+1}}{\text{no. alive at age x}} \tag{16}$$

where q_x denotes age-specific mortality and is a probability.

$$m_x = \frac{\text{no. dying in interval x to x+1}}{\text{no. of years lived x to x+1 by those alive at age x}} \tag{17}$$

where m_x denotes the central death rate, and is a rate.

Age-specific mortality, q_x, and the age-specific central death rate, m_x, are related by the formulae (assuming that average deaths occur half way through the age interval x to x+1):

$$q_x = \frac{m_x}{1 - \frac{1}{2}m_x} \tag{18}$$

$$m_x = \frac{q_x}{1 - \frac{1}{2}q_x} \tag{19}$$

The values of age-specific mortality, q_x, and age-specific central death rate, m_x, are similiar at most ages. However, m_x usually exceeds q_x at very young and very old ages. This is because the within-age distribution of deaths at these extremes tend to be skewed left.[10]

4.3 Continuous Mortality

When one wishes to evaluate the survival and death probabilities when ages and time are real numbers, as opposed to integers, tools other than the life tables are then required.

[10] Carey and Roach [5]

4.3.0.1 The survival function, $S(t)$

is defined as the probability that a newborn will survive beyond time t:

$$S(t) = \mathbb{P}[T_0 > t] \tag{20}$$

where T_0 denotes the random lifetime for a newborn. Returning to the life table,

$$l_x = l_0 S(x) \tag{21}$$

provided that all individuals in the cohort have the same age-pattern of mortality, described by $S(x)$. Thus, the l_x's are proportional to the values which the survival function takes on integer ages x and so the life table can be interpreted as a tabulation of the survival function. Figure 5 illustrates the typical behaviour of the surivial function $S(x)$.

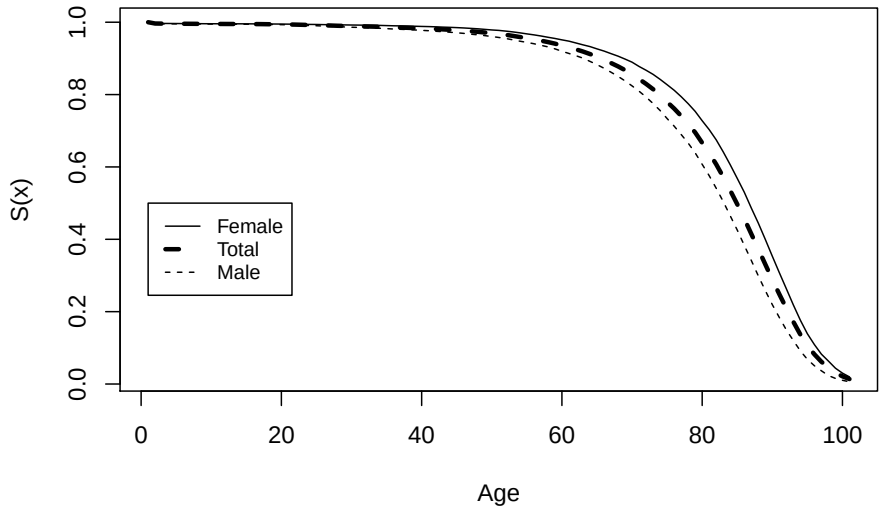


Figure 5: Survival functions for male, female and total populations in Ireland, 2013.

4.3.0.2 The Force of Mortality

Also known as the hazard rate, hazard function or intensity of mortality. Consider the number of survivors at each age x, l_x,[11]

[11] Strictly speaking, the use of the notation l_x is incorrect here, given that l_x is a discrete measure. However, for illustrative purposes it is used to retain a conceptual link with the life table. More correctly, it is defined as: $\mu_x = -\frac{d}{dx} \ln S(x)$. Probabilistically, it can be defined as:

$$\mu_x = \lim_{t \to 0} \frac{\mathbb{P}[T_x \leq t]}{t}. \tag{22}$$

as a continuous function of age x. Then, the force of mortality is defined as the ratio of the rate of decrease of l_x (*i.e.* the instantaneous effect of mortality) at that age to value of l_x.

$$\mu_x = \frac{l_{x+1} - l_x}{l_x} = \frac{-\frac{d}{dx}(l_x)}{l_x} \quad \Longrightarrow \tag{23}$$

$$\mu_x = -\frac{d}{dx}\ln(l_x) \tag{24}$$

where $\frac{d}{dx}$ represents differentiation with respect to age x which indicates the rate of change of l_x over an infintesimally small increment of age. The minus sign is introduced to make μ_x positive since l_x is a decreasing function of age.

Integrating both sides over 0 to n yields:

$$\int_0^n \mu_{x+t} = -\int_0^n \frac{d}{dt}(\ln(l_{x+t}))\,dt = [-\ln(l_{x+t})]_0^n \tag{25}$$

$$= -\ln(l_{x+n}) - \ln(l_x) = -\ln(\frac{l_{x+n}}{l_x}) \tag{26}$$

$$= {}_np_x = exp(-\int_o^n \mu_{x+t}\,dt) \tag{27}$$

If n = 1, this expression can be simplified to:

$$p_x = exp(-\mu_x) \implies \tag{28}$$

$$-\ln(p_x) = \mu_x \tag{29}$$

$$q_x = 1 - exp(-\mu_x) \tag{30}$$

Thus, l_x equals the exponent of the summation of μ_x's from 0 through x.

$$l(x) = l_0[exp(\int_0^x \mu_a\,da)] \tag{31}$$

The force of mortality can be thought of as:

1. A measure of mortality at the precise moment of death
2. The first derivative of the l_x function.

In actuarial modelling, the force of mortality is preferred over age-specific mortality, q_x, because:

1. It is not bounded by unity.
2. It is independent of the size of the age intervals.
3. It forms the argument of numerous parametric mortality functions.

"Although different survivor functions can have teh same basic shape, their hazard functions can differ dramatically... [t]he hazard function is usually more informative about the underlying mechanism of failure than the survivor function. For this reason, modelling the hazard function is an important method for summarising survival data."[12]

4.3.1 The relationship between m_x and μ_x

In a cohort, the central death rate m_x relates to all deaths between the exact age x and the exact age $x+1$. The force of mortality, being a continuous function, gives the instantaneous death rate at each separate age within this interval. More specifically, μ_{x+t} gives the instantaneous death rate at every age $x+t$, where 0<t<1. From one point of view, m_x can be regarded as a kind of average of all the instantaneous death rates in this interval.

It is therefore not surprising to find that m_x is generally quite close to the instantaneous death rate in the middle of that interval. That is to say, there is an approximate relationship

$$m_x \simeq \mu_{x+\frac{1}{2}} \tag{32}$$

A more formal derivation of this approximation is given by Pollard [17].

4.3.2 The relationship between q_x and μ_x

It was previously shown that:

$$1 - q_x = \frac{l_{x+1}}{l_x} \tag{33}$$

There is also a well-known exact equation which states that:

$$\frac{l_{x+1}}{l_x} = e^{-\int_x^{x+1} \mu_x \, dx} \tag{34}$$

Combining these two equations, and assuming that the integral of μ_x between x and x+1 will be close to the value $\mu_{x+\frac{1}{2}}$ at the midpoint of this interval, we obtain the

[12] Tableman and Kim [20].

approximations:

$$\mu_{x+\frac{1}{2}} \simeq -\ln(1-q_x) \tag{35}$$

$$q_x \simeq 1 - e^{-\mu_{x+\frac{1}{2}}} \tag{36}$$

4.3.3 Coninuous Life Expectancy

In the age-continuous context, the life expectancy (or expected lifetime) for a newborn, denoted $\bar{e_0}$, is defined as:

$$\bar{e_0} = \mathbb{E}[T_0] = \int_0^\infty S(t)\,dt \tag{37}$$

(where ∞ in the upper limit of the integral can be replaced by ω, which is functionallly equivalent.) The expected lifetime is often used to compare mortality in various populations.

4.4 Explanatory variables

The mortality metrics described above will generally vary over time as well as age. There could potentially be other explanatory variables, such as frailty, described further on, as well as the more obvious ones, such as gender and nationality. These are termed concomitant variables. Gender and nationality are usually implicitly accounted for and included in life tables, given that life tables are generally constructed for specific states and genders, whilst frailty is termed an unobserved variable that in practice can be difficult to link to observable factors, such as the presence or absence of some type of illness, *etc.*

For the mathematical models considered in this paper, it is assumed, for the short term at least, that the crude death rates will be constant for the next few years. This is a simplifying assumption that, clearly, is not borne out by the improvement in mortality experienced across the world in the last century. Whether such improvements will continue at the same rate or at all into the future and if so, how long into the future, is clearly a question that demographers are faced with but common sense would suggest that there is some asymptotic bound to such improvements in mortality.

5 The Gompertz Law

5.1 Background and derivation

Consider the graph of mortality in Ireland in 2015 plotted against age at Figure 6. It is clear from the graph that there is a distinct regularity in the rate of change with age of mortality. When we plot this rate of change as the ratio of $\frac{q_{x+1}}{q_x}$, over all ages, we see that it is highly irregular from birth to around age 25 to 30 and from about 30 to roughly 90 years of age it is much more constant. In fact, the ratio is about $\frac{q_{x+1}}{q_x} = 1.1$ or 10% change per year.

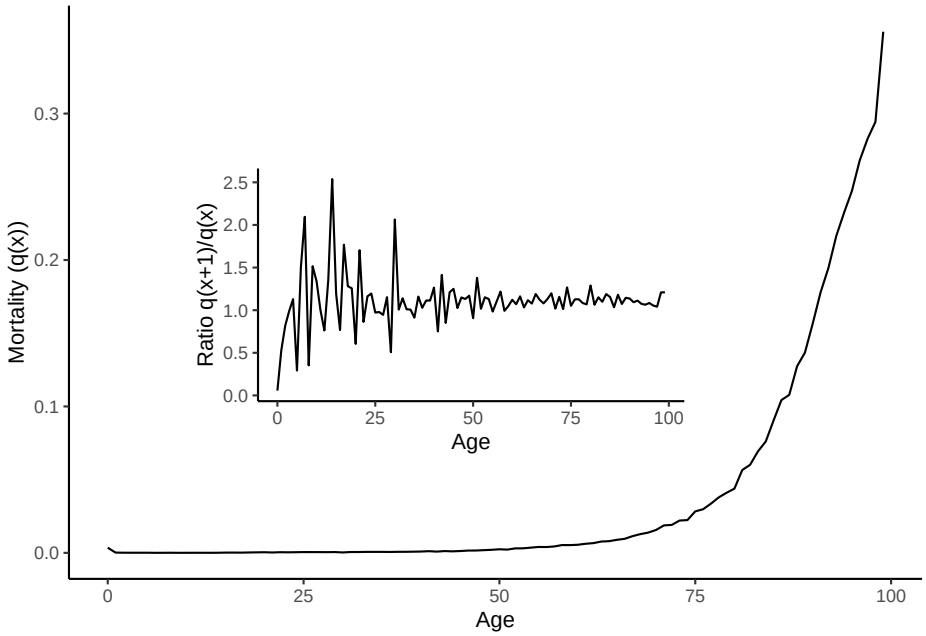


Figure 6: Mortality in Ireland (2015) *Source*: HMD.

Table 1 shows mortality at each age from 70 through 75. The ratios of mortality from 71 to 70, 72 to 71, *etc.* are all close to 1.1, which we will define as λ. In other words, we can predict with a high degree of accuracy the mortality at one age using the product of 1.1 and the mortality at the previous age.

Given that we have identified the constancy of mortality with age, one can derive a mortality model from this information.

Note that since $q_{71} = \lambda q_{70}$ and $q_{72} = \lambda q_{71}$, then by substituting λq_{70} for q_{71}, we find: $q_{72} = \lambda\lambda q_{70}$ or $q_{72} = q_{70}\lambda^2$. By defining q_{70} as the initial level of mortality (denote α), this model can be generalised as:

$$q_x = \alpha\lambda^x \tag{38}$$

Table 1: Ireland (2015) Mortality, q(x) and Rate of Change

Age	q_x	q_{x+1}/q_x
70	0.0156581	1.197775
71	0.0187549	1.018047
72	0.0190934	1.156350
73	0.0220786	1.013020
74	0.0223661	1.267517
75	0.0283494	1.051133

which states that age-specific mortality at age x is the product of the initial level of mortality and the rate of change in mortality raised to the x-th power.

By setting $\lambda = e^{\beta}$, this model can be expressed in continuous form as:

$$\mu_x = \alpha e^{\beta x} \tag{39}$$

where β denotes the exponential parameter. This is also known as the Gompertz parameter, used in one of the most important models in demography - the Gompertz model.

5.2 The Gompertz model

The Gompertz model provides for an exponential increase of mortality with age over a significant span of a lifetime and is commonly regarded as the one of the best models for modeling mortality within demography due to its simplicity and accuracy within that age span. It is also amenable to relatively straightforward parameter estimation techniques. However, as an adjunct of its relative simplicity, is the fact that it does not account for concomitant variables, such as income, social class, *etc.*, which may be regarded as providing further explanatory information on mortality rates.

Benjamin Gompertz[13] proceeded from the idea that the compound growth of human beings can be expressed as continuous growth, at a constant rate, of the instantaneous mortality probability which is given as:

$$\frac{dq(x)}{q(x)} = k\,dx \ ;$$

This results in

$$\ln q(x) = kx + \text{constant}$$

An important assumption of the Gompertz model is that mortality beyond a certain

[13]Gompertz [9]

age (usually around 20) is an exponentially increasing function of age.

The Gompertz model is classically given as:

$$\mu_x = \alpha e^{\beta x}.$$

The model contains two parameters:

1. The initial mortality rate, α, which denotes mortality at the youngest age class in the specified age interval,[14] and;
2. the exponential rate of increase in death rate, β, which denotes the age-specific slope of the mortality function and is sometimes referred to as the Gompertz rate (or alternatively, mortality coefficient or rate of aging). It is a parameter that defines the rate of change of mortality with age.

This model can be expressed in a linearised version as:

$$\ln \mu_x = \alpha + \beta x. \tag{40}$$

The above form of the Gompertz model derives the force of mortality as a function of age. The Gompertz model can also be specified in terms of the cohort survival function, l_x, where:

$$l_x = e^{[(\frac{\alpha}{\beta})(1-e^{\beta x})]} \tag{41}$$

In addition, the Gompertz model provides two useful formulae:

- The mortality doubling time, denoted MDT, defined as the time required for the mortality to increase two-fold (MDT $= \frac{\ln(2)}{\beta}$), and;
- The estimated maximum life span, denoted T_{max}, and defined as the age when a population subject to Gompertzian mortality rates has diminished to one survivor ($T_{max} = \frac{\ln[1+\frac{\beta \ln(N)}{\alpha}]}{\beta}$).

6 Heterogeneity in Human Mortality

An assumption with many mortality models, such as the Gompertz equation, is that all individuals at birth experience the same probability of survival at each age. This assumption is unrealistic. As populations age, the composition changes. Over time, members of sub-cohorts with higher death rates die out faster. Because the curve of the force of mortality, μ_x, for humans cannot be adequately modelled by a single

[14] Usually age 0 in most demographic studies.

Gompertz function, due to the fact that it exhibits different characteristics over different age ranges, one needs to consider how this heterogeneity can be incorporated into mathematical models and accounted for more generally.

Kirkwood[15] has this to say on the phenomenon:

> "the phenomenon of the apparent plateau in late-life mortality requires explanation. One attractive solution to this enigma is population heterogeneity. If the population comprises a set of individuals who differ in the inherent robustness or rate of ageing, the frailest individuals tend to die soonest, leaving the most robust individuals to predominate at the oldest ages. In such a scenario, it is possible for a late-life mortality plateau, or even a decline, to be seen at the population level, even if the various individuals that comprise the population all experience an increasing probability of dying with advancing age."

Pitacco *et al*[16] state that:

> "...it has been observed that the force of mortality is slowly increasing at very old ages, approaching a rather flat shape. In other words, the exponential rate of mortality increase at very old ages is not constant, as for example, in Gompertz's law, but declines ... as classical mortality laws may fail in representing the very old-age mortality, shifting from the exponential assumption may be necessary in order to fit the relevant pattern of mortality."

The authors then proceed to refer to several alternative models that have been proposed to deal with this phenomenon, such as the Heligman-Pollard family of laws and the Perks' laws to deal with mortality at highest ages.

6.1 The Concept of Frailty

Demographic heterogeneity is the concept of having sub-groups within a population which are endowed with different levels of frailty. In frailty models, a distinction is made between the mortality schedules for individuals and the entire population.[17] Individual frailty is based on the consideration that each single individual in a population has their own specific traits and faces certain mortality dynamics which are reflected in the chances of survival for that individual due, for example, to genetics, exposure to risks such as smoking, dangerous work practices, *etc.* These individuals can then be

[15] Kirkwood [13]
[16] Pitacco [15]
[17] Vaupel and Missov [22].

aggregated into more or less homogeneous sub-groups, with individuals within each sub-group sharing, to some extent at least, those frailty characteristics.

James Vaupel is most closely associated with developing such heterogeneous models based on the concept of frailty. In the frailty model developed by Vaupel *et al*[18], an individual's frailty is a relative-risk that is fixed for life.

The frailty parameter, z, is a random variable that is unobserved and is defined according to the relationship:

$$\mu(x, y, z) = z \cdot \mu(x, y, 1) \tag{42}$$

where an individual with the frailty parameter $z = 1$ is a 'standard' individual. This means that an individual with a frailty of 2 is twice as likely to die at any particular age, x, and time, y, and an individual with a frailty of $\frac{1}{2}$ is half as likely to die. For simplicity, the subscript and arguments are often dropped so the previous equation is reduced to: $\mu(z) = z \cdot \mu$.

Let s be the probability that an individual will survive to age x and H denote the cumulative hazard function through age x. Then:

$$s = e^{-H} \quad \text{and it follows that:} \tag{43}$$

$$s(z) = s^z \ (= \text{survival}) \tag{44}$$

Vaupel *et al*[19] assumed that frailty at birth is gamma distributed:

$$f_0(z) = \lambda^k \cdot z^{k-1} \cdot \frac{e^{-\lambda z}}{\Gamma(k)} \tag{45}$$

where λ and k are the parameters of the distribution and $\Gamma(k)$ is the gamma distribution. They chose this distribution as it is analytically tractable and readily computable. They note that it is a flexible distribution that takes on a variety of shapes as k varies.

6.2 Risk Factors and Modelling methodologies

The risk factors that potentially impinge upon an individual's (or sub-cohort's) frailty can be divided into those that are described as 'observable' and those described as 'un-observable'. In the former category would be such factors as age, gender, *etc.* and in the latter, such factors as congenital characteristics, lifestyle, *etc.*

Pitacco[20] outlined a survey of research and scientific contributions on heterogeneity

[18]Vaupel, Manton, and Stallard [21]
[19]Vaupel, Manton, and Stallard [21]
[20]Pitacco [14]

in mortality. A summary of the classification of such heterogeneous models is shown at Figure 7.

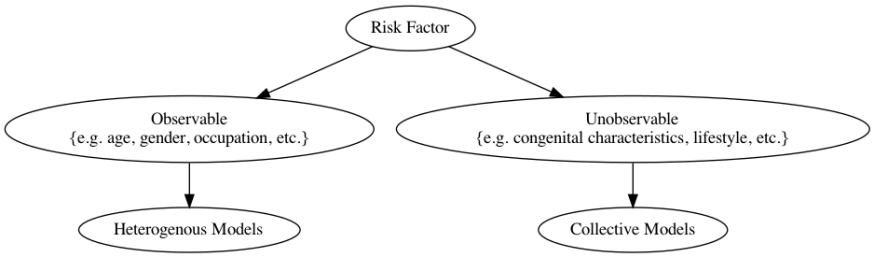


Figure 7: Typology of mortality models depending on type of risk factor.

The authors state that "when allowing for unobservable heterogeneity factors, two approaches can be adopted:

- a discrete approach, according to which heterogeneity is expressed through a (finite) mixture of appropriate functions
- a continuous approach, based on a non-negative real valued variable, called the frailty, whose role is to include all unobservable factors influencing the individual mortality."

6.2.1 Hypotheses for differing mortality rates

Various hypothetical explanations have been profferred for differing mortality rates across sub-populations, ranging from genetic to environmental explanations. Avraam et al. [3] make an assumption that 'responses to environmental factors are largely shaped by an organism's genetic landscape and... can affect the dynamics of ageing and mortality.' However, such an assumption did not influence the formulation of their model and the authors merely attempted to ascertain whether such an assumption was contradicted by the data or not. No such assumptions concerning what the underlying causative effects of differing mortality rate profiles might be are made in this paper.

Pitacco [15] states that:

> "As regards observable factors, mortality depends on: (1) biological and physiological factors, such as age, gender, genotype; (2) features of the living environment; in particular: climate and pollution, nutritional standards (mainly with reference to excesses and deficiencies in diet), population density, hygienic and sanitary conditions; (3) occupation, in particular in relation to professional disabilities or exposure to injury, and educational attainment; (4) individual lifestyle, in particular with regard to nutrition, alcohol and drug consumption, smoking, physical activities and pastimes;

(5) current health conditions, personal and/or family medical history, civil status, and so on.

The author continues:

> Heterogeneity of a population in respect of mortality can be explained by differences among the individuals; some of these are observable, as discussed in the previous section, whilst others (e.g. the individual's attitude towards health, some congenital personal characteristics) are unobservable.

The author then proceeds to only deal with the latter, continuous, approach.

6.2.2 Issues with existing models

One of the main issues with existing attempts to model mortality which account for heterogeneity in human populations is their complexity. Moreover, the fit of such models is less than optimal.

7 The Heterogeneous Gompertz Law

One of the shortcomings of the standard Gompertz model is that it is valid only over a limited age cohort or range, commonly regarded as being between around 20 and 60 and does not capture the dynamics of mortality outside those ages. Mortality rates predicted by a Gompertz model, whose parameters have been designed or fitted for that age cohort will not fit actual mortality rates for very young and very old age cohorts outside that middle age cohort. This is due to several factors, such as high infant mortality rates and a phenomenon whereby there is a flattening out of the death rate at older ages. In particular, the "problem of computing μ_x for $0 \leq x \leq 2$ is extremely complex"[21]

The behaviour of the force of mortality as a function may roughly be split into three types for three age ranges[22]:

1. The age range of *c.* 0 - 10 years is a period of rapidly declining mortality rates
2. In the age range from 10 to 30, μ_x is an increasing function, levelling off to a plateau between 20 and 30.
3. The remaining years show an exponential increase in the force of mortality.

[21] Pollard [17]. The author states that "in the Australian life table (males) for 1961, for example, μ_x is only given for ages greater than two years. Approximate values can be obtained by fitting a hyperbola to the l_x function".

[22] Atkinson and Dickson [1]

7.0.1 Discrete version of the Gompertz model

According to the discrete version of the Gompertz law, the central death rate (or mortality rate) at age i is represented as:

$$m_i = m_0 e^{\beta i}$$

where m_0 is the initial mortality at age $i = 0$ and is the discrete analogue of the force of mortality, μ_x for a continuous model. This is an age-dependent, intrinsic model in that the predicted mortality rate depends only on the parameters m_0 and β and on the age of the individual, i, within the cohort under consideration.

7.1 Discrete mathematical model of heterogeneous populations

As outlined above, Avraam, Magalhaes, and Vasiev [2] presented a mathematical model to predict mortality rates based on an additive gompertz model of two or more heterogeneous populations. The model is discrete, rather than continuous, in that a mortality rate is predicted for each discrete age from i = 0, 1, 2,. . . ,ω, where ω is the upper age limit of the population under consideration.

The model is predictive, rather than explanatory in that it does not make any *a priori* assumptions as to what the underlying processes generating the mortality patterns may be. The primary criteria in generating the model is that it fits the actual data well.

The authors do make some tentative suggestions as to what the underlying processes may be, mainly on genetic grounds, but make it clear that exploring such causative processes would be for a future study. In their work, the authors aimed to model mortality across the whole lifespan presuming that the rate of mortality changes over age according to the Gompertz law. They state:

> "Although many other models have been used to describe mortality dynamics over age Pletcher [16] there is a genuine feeling that the fundamental processes underlying mortality should result in exponential law Yashin, Iachine, and Begun [23]. Therefore we analyse whether the deviations from Gompertz law can be explained by the heterogeneity of populations while the mortality in each sub-population is still described by the Gompertz law."

The model is specified, in the discrete case, as follows:

$$m_i = \frac{\sum_{j=1}^{n} \frac{\rho_{ji} m_{j0} e^{\beta_j i}}{1 + 0.5 m_{j0} e^{\beta_j i}}}{1 - 0.5 \sum_{j=1}^{n} \frac{\rho_{ji} m_{j0} e^{\beta_j i}}{1 + 0.5 m_{j0} e^{\beta_j i}}} \quad \text{where:}(\#eq : het_e q) \tag{46}$$

$$\rho_{ji} = \frac{\rho_{j0} \prod_{k=0}^{i-1} \left(\frac{1 - 0.5 m_{j0} e^{\beta_j k}}{1 + 0.5 m_{j0} e^{\beta_j k}} \right)}{\sum_{s=1}^{n} \left(\rho_{s0} \prod_{k=0}^{i-1} \left(\frac{1 - 0.5 m_{s0} e^{\beta_s k}}{1 + 0.5 m_{s0} e^{\beta_s k}} \right) \right)} \tag{47}$$

Using these formulae, the authors did a simulation analysis and constructed plots, shown at Figure 8, to illustrate the effect of varying the various parameters, $\rho_{j0}, m_{j0}, \beta_j$.

To this extent, the heterogeneous Gompertz model addresses one of the shortcomings of the standard Gompertz model, *i.e.* its relative simplicity which cannot account for the homogeneity in mortality rates that is observed to occur naturally in large populations. Other mathematical models do not explicitly attempt to account for this in cohort form but rather by way of parameterisation. The precise meanings of the parameters in such models thus tend to become obscured, unlike the straightforward meaning of the parameters in the heterogeneous Gompertz model, which represent the characteristics of specific cohorts of the population. Having said that, the underlying causes for such heterogeneity amongst the cohorts is not explicitly addressed, as mentioned, by the heterogeneous Gompertz model.

The discrete formulae for constructing the heterogeneous Gompertz models described by the authors was implemented independently in R by this author and an R package was created, containing same.[23] The functions in that package were then used to reproduce the simulation done by Avraam, Magalhaes, and Vasiev [2] and the plots of the simulated data with the varying parameters were reproduced using R.

7.1.1 Models applied to Swedish and US data

Avraam, Magalhaes, and Vasiev [2] then proceeded to fit data from the Human Mortality Database[24] for Sweden to the model. Heterogeneous populations were created from different combinations of sub-populations. Using least squares to fit the data, they found that the best model was composed of four sub-populations with the following parameters:

[23] It can be found on github at https://github.com/pachristopher/projectRpkg.
[24] *Human Mortality Database* [11]

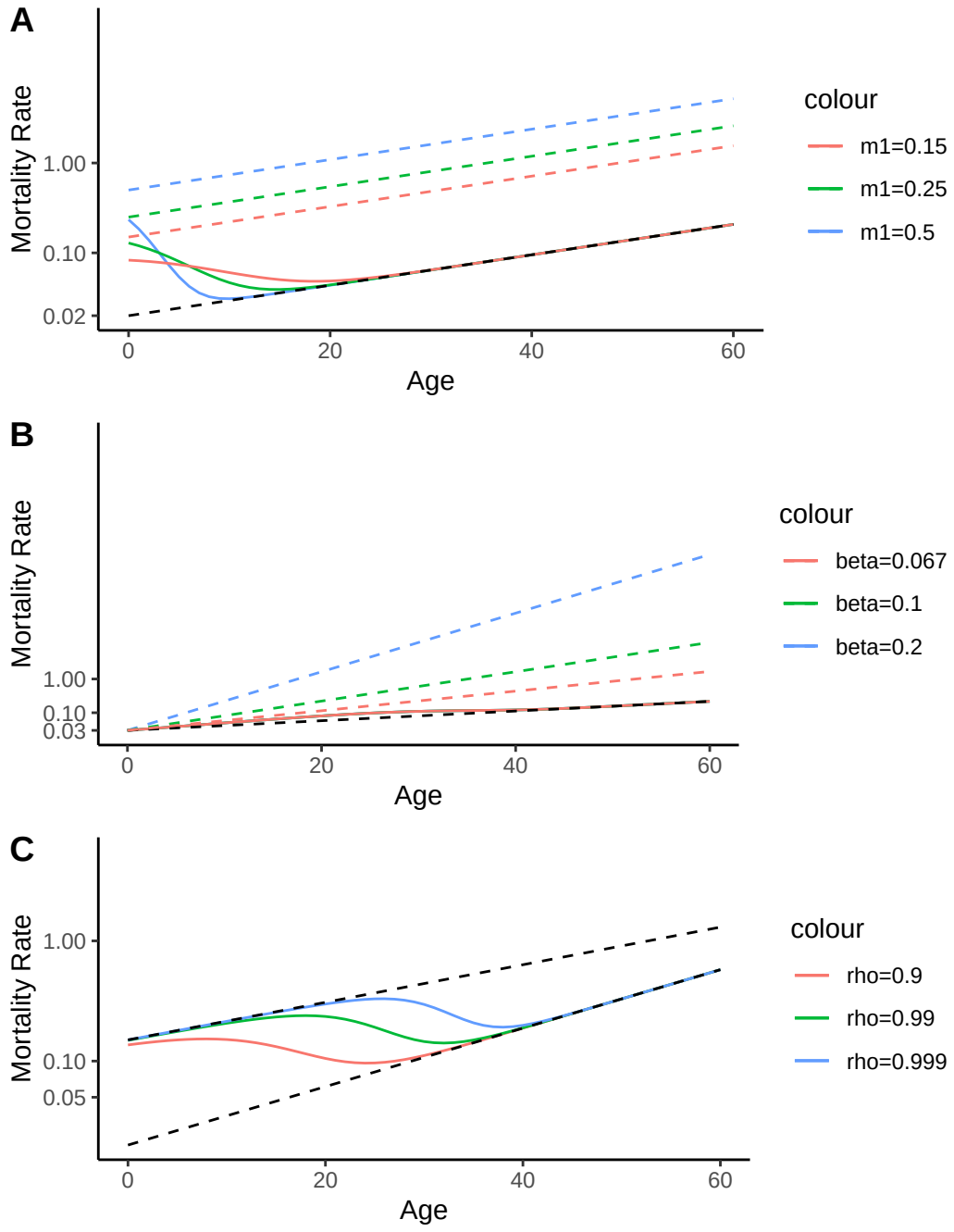


Figure 8: The effect of varying model parameters on the mortality dynamics of a heterogeneous population consisting of two subpopulations.
A: The effect of varying the initial mortality rate for one of the subpopulations. Subpopulations have equal initial sizes and equal ageing slopes. The total mortality of the entire population is represented by a solid line with the colour of the corresponding dashed line.
B: The effect of varying the ageing slope. Subpopulations have equal initial sizes $\rho_{10} = \rho_{20} = 0.5$ and equal initial mortality rates m10 = m20 = 0.03. The rate of ageing β_1 takes the values 0.2, 0.1 and 0.067 for the blue, green and red dashed lines, respectively, while $\beta_2 = 0.033$ is constant. The total mortality of the entire population is represented by a solid line with the colour of the corresponding dashed line (indicating the value of β_1).
C: The effect of varying the initial size of the subpopulation. Two subpopulations (dashed lines) with different ageing slopes ($\beta_1 = 0.036, \beta_2 = 0.056$) and different initial mortality rates (m10 = 0.15, m20 = 0.02) are considered. Blue, green and red lines show the total mortality of a whole population where the initial fraction ρ_{10} is 0.9, 0.99 and 0.999 correspondingly.

- 1st sub-population: $m_{10} = 1.6139, \rho_{10} = 0.00266, \beta_1 = 0.67e-5$;
- 2nd sub-population: $m_{10} = 0.108, \rho_{10} = 0.00057, \beta_1 = 0.2685$;
- 3rd sub-population: $m_{10} = 0.00052, \rho_{10} = 0.00460, \beta_1 = 0.2558$;
- 4th sub-population: $m_{10} = 0.000013146, \rho_{10} = 0.99217, \beta_1 = 0.1041$.

Fitting that model to the Swedish mortality rate data for 2007, they found a sum of squared residuals of 3.229884 and a BIC -336.3785.

A plot of the line fitted by Avraam, Magalhaes, and Vasiev [2] with the data points overlaid is shown at Figure 9.

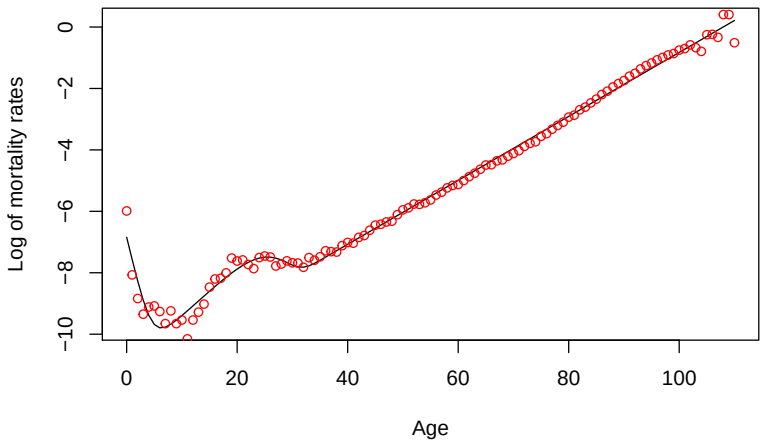


Figure 9: The modelled mortality of the heterogeneous population is given by the solid line and the Swedish mortality rate data from HMD is denoted by the red dots.

8 Estimating Parameters of the Heterogeneous Gompertz Model

A heterogeneous Gompertz model was specified as per 46 above. To find values for the free parameters, ρ_{0i}, β_i and m_{0i}, that could minimise the sum of squared residuals (being the difference between the theoretical prediction and the observation data) and therefore to fit the data, the least squares (LS) method was used. This method was implemented using the Solver tool in LibreOffice Calc.[25]

[25]The specific solver engine used was the non-linear DEPS (Differential Evolution Particle Swarm) Evolutionary Algorithm. See: https://help.libreoffice.org/latest/ug/text/scalc/01/solver_options_algo.html. Some background on this algorithm can be found in Zhang, Wang, and Ji [24], *Evolutionary and swarm intelligence algorithms* [8] and Rooy [18].

Table 2: Parameters for 4 sub-population heterogeneous Gompertz model fitted to 2007 Swedish mortality data from HMD.

Parameters	Values	Avraams's values
m_{10}	1.4682286	1.6139000
β_1	-0.6000526	0.0000067
ρ_{10}	0.0028923	0.0026600
m_{20}	0.0824330	0.1080000
β_2	0.3525918	0.2685000
ρ_{20}	0.0003602	0.0005700
m_{30}	0.0011760	0.0005200
β_3	0.2066538	0.2558000
ρ_{30}	0.0047175	0.0046000
m_{40}	0.0000132	0.0000130
β_4	0.1034126	0.1041000
ρ_{40}	0.9920301	0.9921700

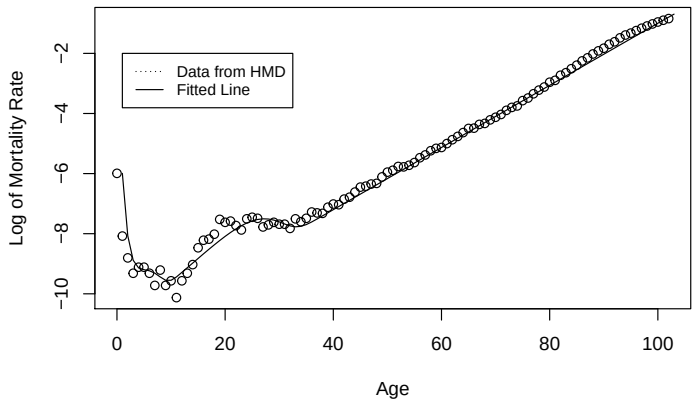
8.0.1 Fitting Swedish data

As a preliminary calibration, parameters were fitted to the four sub-population model using 2007 Swedish data from HMD, as used by Avraam, Magalhaes, and Vasiev [2], using the LS procedure described above to see how closely they conformed to the results obtained by Avraam *et al.* The parameters obtained from fitting the model to the 2007 Swedish mortality data using LibreOffice solver tool are shown at Table 2. Also shown for comparison are the parameter values obtained by Avraam, Magalhaes, and Vasiev [2].

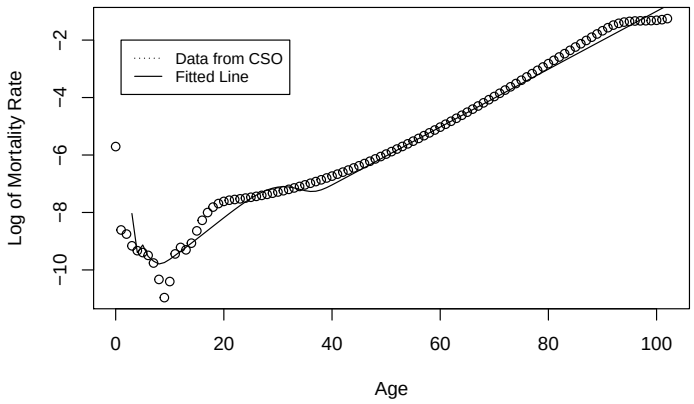
The BIC obtained was -319.7 and the residual sum of squares was 2.817. The values of the parameters and the goodness of fit measures approximate quite closely those obtained by Avraam, Magalhaes, and Vasiev [2], where the BIC obtained was -336.3 and the RSS was 3.2. This is encouraging, given that the method implemented by Avraam, Magalhaes, and Vasiev [2] used a different nonlinear regression algorithm (provided by the command DataFit in Maple which is included in the DirectSearch package).

Plots of the Swedish 2007 mortality data used by Avraam, Magalhaes, and Vasiev [2] with the fitted line from the 4 sub-population heterogeneous Gompertz model superimposed and the Irish mortality data from the CSO 2015-2017 life table, again with the four sub-population heterogeneous Gompertz model fitted line superimposed are shown at Figure 10.

```
[1] 102
```



(a) Swedish 2007 data from HMD



(b) Irish 2015-2017 data from CSO

Figure 10: Plots of the 2007 Swedish and 2015-2017 Irish mortality data with fitted lines from the 4 sub-population heterogeneous Gompertz model superimposed

8.0.2 Fitting Irish CSO data

Using the LS method, parameters of the models for heterogeneous populations consisting of three sub-populations with data from the 2015-2017 life table from the Central Statistics Office of Ireland (CSO) were fitted. A model consisting of four sub-populations was then fitted to the same data. The next task is to find out which of these two models is a better fit. The criterion used for evaluation of how well the model fits the data is the Bayesian Information Criterion (BIC) (Schwarz [19]) which is given by the formula:

$$\text{BIC} \ = n_d \ln(\hat{\sigma_e^2}) + k \ln(n_d) \tag{48}$$

where n_d is the number of data points, $\hat{\sigma_e^2}$ is the sum of squared residuals divided by the number of data points and k is the number of free parameters. The model with the lower value of BIC implies a better fit to the data. According to the BIC, the heterogeneous model with four sub-populations (BIC = -232.62) fits the CSO data better than the model with three sub-populations (BIC = -176.12). Plots for the three and four sub-population models, as fitted to the CSO data, are at Figure 11 and Figure 12.

Table 3 and Table 4 show the parameters for the three and four sub-population models fitted to the Irish CSO data, respectively.

Table 3: Parameters for 3 sub-population heterogeneous Gompertz model fitted to male 2015-2017 CSO data.

ρ_{0i}		m_{0i}		β_i	
ρ_{01}	0.002237328077671	m_{01}	4.54973908257018	β_1	-0.1500761318802
ρ_{02}	0.000557174246818	m_{02}	0.368200387234967	β_2	0.211537687783045
ρ_{03}	0.997205497675511	m_{03}	3.568107751643e-05	β_3	0.09139640495285

Table 4: Parameters for 4 sub-population heterogeneous Gompertz model fitted to 2015-2017 CSO data.

ρ_{0i}		m_{0i}		β_i	
ρ_{01}	0.002547208824796	m_{01}	3.20393321656011	β_1	0.010226506161476
ρ_{02}	0.000650784396633	m_{02}	0.282271320800018	β_2	0.221852537463537
ρ_{03}	0.006435402652686	m_{03}	0.000570813106869	β_3	0.206653841941924
ρ_{04}	0.990366604125885	m_{04}	1.717315579e-05	β_4	0.100761335928826

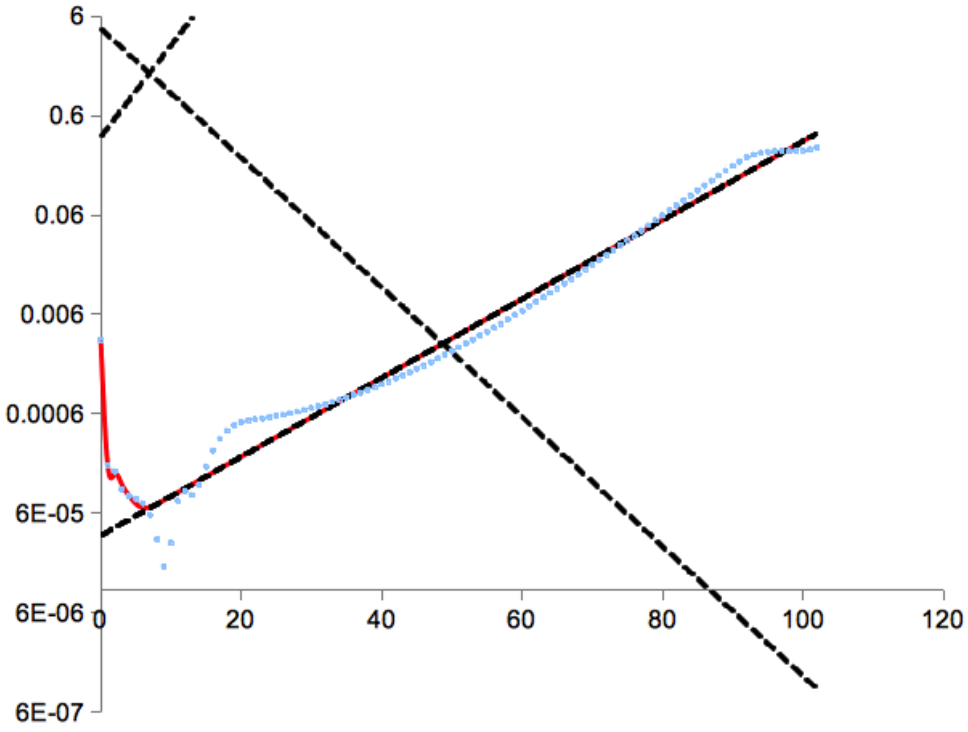


Figure 11: Fitting the heterogeneous model to 2015-2017 Irish mortality data from the CSO using Least Squares. The data is denoted by blue points, mortality rates of the model sub-populations by dashed lines and the mortality rates of the whole population by the red solid curve (Three sub-population model).

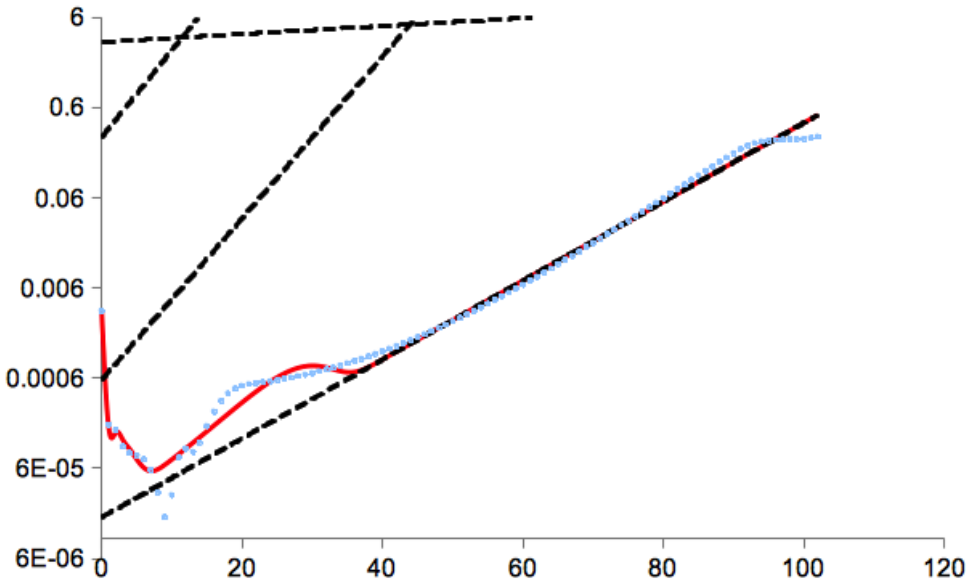


Figure 12: Fitting the heterogeneous model to 2015-2017 Irish mortality data from the CSO using Least Squares. The data is denoted by blue points, mortality rates of the model sub-populations by dashed lines and the mortality rates of the whole population by the red solid curve. (Four sub-population model)

8.0.3 Irish Life Insurance data

The same exercise was performed using different mortality data provided by the Irish Life insurance co. and the results are set out below. Plots for the three and four sub-population models, as fitted to the Irish Life Insurance Co. data, are at Figure 13.

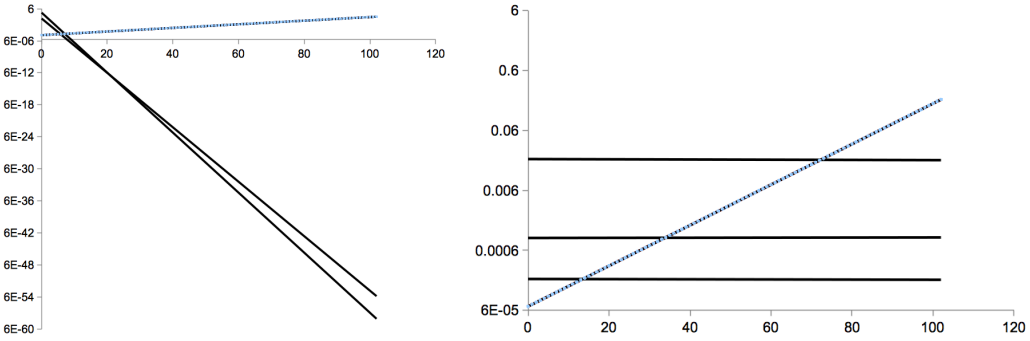
As can be seen, the three and four sub-population heterogeneous Gompertz models are unnecessarily complex for the Irish Life insurance data. This is not surprising, given the relative homogeneity within the cohort captured by that dataset. Accordingly, the basic Gompertz model may be a better way to model mortality for this cohort of Irish Life policyholders. The methodology for doing this is outlined hereafter.

8.1 Modelling the Irish Life Insurance Data with a Simple Homogeneous Gompertz Model

A spreadsheet of Irish Life insurance data, containing the age in one column and the mortality rate for different years in the other columns was read into R.

A log transformation was done on the mortality rates. A linear model of the following form:

$$\ln(q_i) = \alpha + \beta i$$



(a) Three sub-population model (b) Four sub-population model

Figure 13: Fitting the heterogeneous model to 2015 Irish mortality data from the Irish Life Insurance Co. using Least Squares. The data is denoted by blue points, mortality rates of the model sub-populations by dashed lines and the mortality rates of the whole population by the red solid curve.

was then fitted using the `lm()` function in base R, where q_i represents the mortality rate at age i, or the probability that a person of age i will die before reaching age $i+1$.

This equation can be transformed using the exponential function as follows:

$$\ln(q_i) = \alpha + \beta i \implies \tag{49}$$

$$q_i = Ae^{\beta i} \quad \text{where; } A = e^{\alpha} \tag{50}$$

to give a form that is more similiar to that used to describe the Gompertz Law in most texts.

$$\mu(x) = \alpha e^{\beta x}$$

As a preliminary step, the log-transformed mortality rates for the period 2015 were modelled using a linear regression model of the form described above. The model and its output is shown below:

```
Call:
lm(formula = data2$X2015_ln ~ data2$Age)

Residuals:
       Min          1Q      Median          3Q         Max
-3.558e-04 -2.560e-06  3.030e-06  4.220e-06  7.116e-04
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.572e+00  2.038e-05 -469582   <2e-16 ***
data2$Age    7.793e-02  3.202e-07  243350   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0001081 on 109 degrees of freedom
Multiple R-squared:      1, Adjusted R-squared:      1
F-statistic: 5.922e+10 on 1 and 109 DF,  p-value: < 2.2e-16
```

Based on the output from this linear model, the equation for the model based on the 2015 data (and similarly for the data from the other years) can be written as:

$$ln(q_i) = -9.752 + 0.078 * i \implies$$
$$q_i = 6.956 \times 10^{-5} e^{0.078 * i}$$

8.1.1 Fitting Gompertz model to Irish insurance data

The following code fits the linearised Gompertz model to the remainder of the Irish insurance data and analyses how the two parameters change over time for each cohort.

```
# Select columns from data2 if their names start with "expression"
loop.vector <- names(data2[, grep(pattern="X[0-9]{4}_ln", colnames(data2))])
lm.test <- vector("list", length(loop.vector)) #create empty list to store models
for(i in seq_along(loop.vector)){
  lm.test[[i]] <- lm(reformulate("Age", loop.vector[i]), data = data2)
}

cfs <- lapply(lm.test, coef) # list of model coefficients
names(cfs) <- loop.vector # give them names
cfs <- t(as.data.frame(cfs))
rownames(cfs) <- c('2015','2005','1995','1985','1975','1965','1955','1945','1935')
knitr::kable(cfs, caption = "Evolution of parameters of Gompertz model for Irish Life ins
```

As can be seen from Table 5, the intercept term has been steadily decreasing from 1935 to the present. Whereas the slope coefficient has varied around a fairly steady average. Concretely, this translates into a steadily decreasing mortality rate over all

Table 5: Evolution of parameters of Gompertz model for Irish Life insurance data.

	(Intercept)	Age
2015	-9.572022	0.0779305
2005	-9.495081	0.0791065
1995	-9.176914	0.0765957
1985	-8.981957	0.0777089
1975	-8.894596	0.0774555
1965	-8.672472	0.0769277
1955	-8.626314	0.0788510
1945	-8.360872	0.0778124
1935	-8.150926	0.0780908

ages over the period with a fairly constant aging rate. Consistent with this analysis, the lines are roughly parallel, with lower intercepts as the years progress.

8.1.2 Plotting the Irish data

As the next step, plots of the fitted lines were produced for each period in the Irish insurance mortality dataset, *i.e.* from 1935 to 2015. The fitted lines from the above model paramaters are drawn on the same plot in different colours. They are shown at Figure 14.

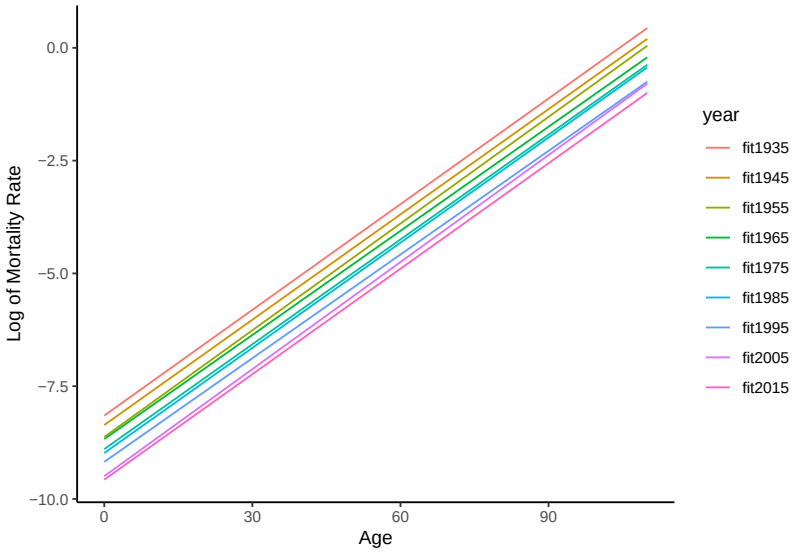


Figure 14: Plots of Age vs. Log of Mortality rate for Ireland.

When plotting the Irish Life insurance data, it is striking how well the model fits, *i.e.* how 'straight' the plot line is and how well the fitted line 'fits' it. This was in contrast to the plots of the Irish mortality rates from the CSO and HMD data, which

displayed the characteristic 'dip' in the infant years, a feature which was absent from the Irish Life insurance data.

A reasonable initial hypothesis may be that the Irish Life insurance data is not representative of the whole Irish population given that it is data on life policy holders only. In that sense, it may have been 'cleansed' or 'self-selected', *i.e.* it did not include the more risky cohorts from the population as a whole and was thus more homogeneous than mortality data from the CSO or HMD, which represents the whole population of the country, rather than a subset of it.

8.2 Life Expectancy

Each of these models can be used to calculate the life expectancy at birth, referred to as $T(0)$ in the actuarial literature (Atkinson and Dickson [1]). Briefly, the calculation involved obtaining p_x, the probability that a person aged x will survive to age $x+1$, by subtracting q_x, the probability that a person aged x will die before reaching age $x+1$, from 1. $T(x)$, the length of the future lifetime of the individual now aged exactly x, is then obtained by summing p_x over all years from the present age, x, to ω, the upper age limit of the population:

$$T(x) = \sum_{i=x}^{\omega} p_i \implies \tag{51}$$

$$T(0) = \sum_{i=0}^{110} p_i \tag{52}$$

$$T(0) = \sum_{i=0}^{110} (1 - q_i) \tag{53}$$

$$T(0) = \sum_{i=0}^{110} (1 - e^{\alpha} e^{\beta i}) \tag{54}$$

given that: $\omega = 110$.

The code below calculates the theoretical life expectancy at birth for each of the fitted models and they are printed out at Table 6.

```
age_vec <- 0:110 # create vector of Ages: 0-110
# create empty vector to store life expectancy at birth
life_exp <- vector("list", length(lm.test))
for(i in seq_along(lm.test)){
  ln_q_hat <- predict(lm.test[[i]], Age = age_vec)
  p1 <- 1
  p_hat <- Reduce(function(v, x) v*(1-x), x=exp(ln_q_hat), init=p1, accumulate=TRUE)
```

Table 6: Theoretical life expectancy calculated from Gompertz model for Irish insurance data between 1935 and 2015.

1935	1945	1955	1965	1975	1985	1995	2005	2015
64.36	67.2	69.8	71.82	74.26	75.17	78.6	80.5	82.5

Table 7: Actual life expectancy in Ireland between 1955 and 2015 from HMD data.

	1955	1965	1975	1985	1995	2005	2015
	68	70.64	71.7	73.56	75.48	78.92	81.43

```
  life_exp[[i]] <- sum(p_hat[2:length(p_hat)])
}
names(life_exp) <- loop.vector # add names to elements of the vector
th_life_exp_df <- as.data.frame(life_exp)
colnames(th_life_exp_df) <- c('2015','2005','1995','1985','1975','1965','1955','1945','19
knitr::kable(round(rev(th_life_exp_df),2), caption = "Theoretical life expectancy calcula
```

The life expectancy figures look quite high, although perhaps this would not be surprising if the insurance data reflected the fact that the cohort captured by the data was healthier, on average, than the general population. It is also important to note that life expectancy will be naturally over-estimated when simulated from the Gompertz model.

It would be instructive to see a similiar table for the general Irish population as opposed to one that had been screened by life insurers. A further data source is the Human Mortality Database (HMD). This data can be accessed directly with the `demography` package. A table, similar to the table for the fitted life expectancies from the linear model can also be produced from the HMD data. This is at Table 7. As the HMD only provides data from 1950 onwards, its range is more truncated than the Irish insurance data, beginning at 1955, rather than 1935.

A further method of investigating the life expectancy was made easier now that the CSO provides an R package (`csodata`) to allow direct importation of datasets into R. This was used to import the period Life Expectancy at Various Ages (VSA30) dataset held by the CSO and plot a graph from it. This data is split by gender, unlike the insurance data.

Figure 15 shows how the average life expectancy for both males and females has been steadily improving in Ireland from about 50 for both sexes in 1871 to about 83 for females and just under 80 for males in 2016. That is nearly a doubling of life expectancy over a period of about 150 years. Various explanations exist for this phenomenal growth in life expectancy, from medicinal and hygiene improvements, to

improvements in air quality and so forth. There is some controversy as to whether, and for how long, improvements in life expectancy can continue into the future. Whilst there is undoubtedly a limit to the extent to which life expectancy can continue to improve, rather the debate centres around what that limit may be and when it will be achieved. Already, some developed countries are reporting a drop in life expectancy amongst certain cohorts of their populations. Ho and Hendi [10] report that "the United Kingdom and the United States appear to be experiencing stagnating or continued declines in life expectancy, raising questions about future trends in these countries."

It is obvious that the life expectancy figures are lower than for the insurance data, particularly for the earlier years.

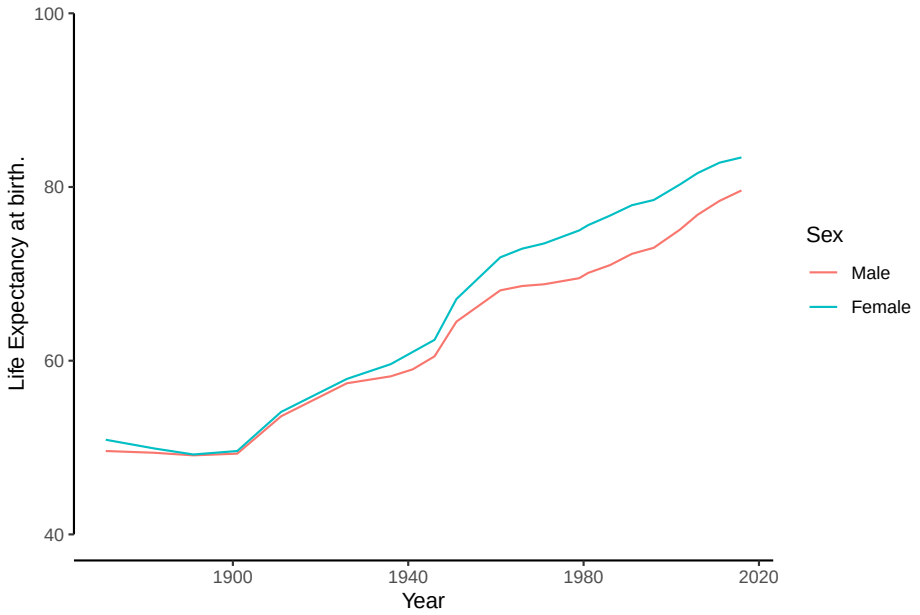


Figure 15: Plot of life expectancy at birth: Ireland: 1871-2016

A plot of Age vs. Life expectancy at that age, across the period from 1950 to 2017 is shown at Figure 16, using the HMD data. The rainbow colour scheme shows the older years represented by red lines and more recent years by the colours at the violet end of the visible spectrum.

An interesting feature that emerges from this plot, apart from the increasing life expectancy across all ages over the period, is that the high infant mortality apparent in the 1950s and 1960s (as represented by the kinks in the red curves, representing a curious feature whereby a two-year old could expect to live longer than an infant aged less than one, despite already being over a year older) has all but disappeared post-2000.

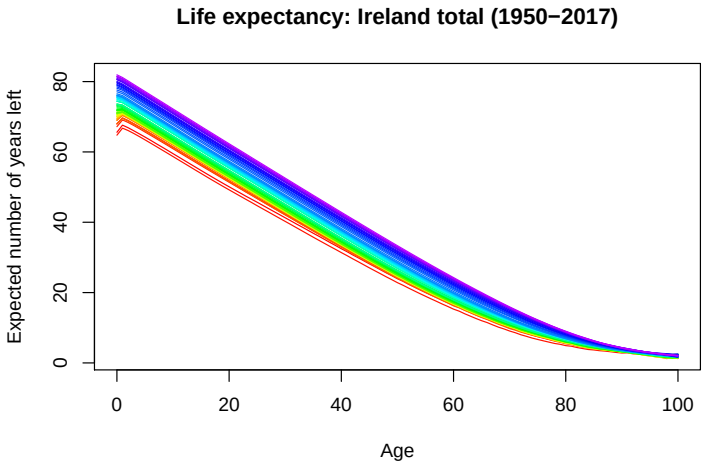


Figure 16: Change in life expectancy across all ages in Ireland between 1950 and 2017.

9 Simulation of Age at Death

Another way to estimate life expectancy is to simulate the age at death for a sufficiently large number of hypothetical lives. The simulation can then be cross-checked to see how closely it matches the theoretical life expectancy obtained from the mathematical model. The advantage of simulation is that the standard deviation of the average life expectancy can be obtained and a histogram of the underlying age at death distribution can be plotted.

R code was implemented on the CSO life table for 2015-2017 in the LifeExp.xlsx file. A histogram of the simulated ages at death is shown at Figure 17.

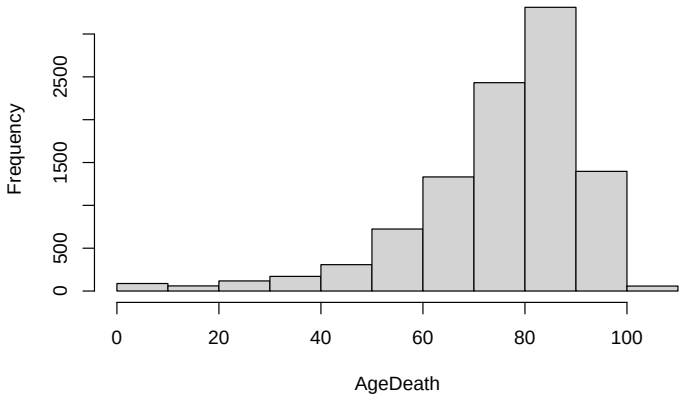


Figure 17: Histogram of Simulated Age at Death.

The mean simulated age at death is 75.7411 and the standard deviation is 16.8565621. The histogram, unsurprisingly, has a left-skew. This fits quite closely to the theoretical Life expectancy at age zero, $T(0) = 75.8$, which was calculated from the life table as follows:

$$\mathring{e}_x = \frac{1}{2}L_0 + \sum_{x=1}^{\omega}(L_x) \tag{55}$$

$$\text{where } L_x = \prod_{i=0}^{x}(1 - q_i) \tag{56}$$

A plot of the theoretical life expectancy calculated from the life table is at Figure 18.

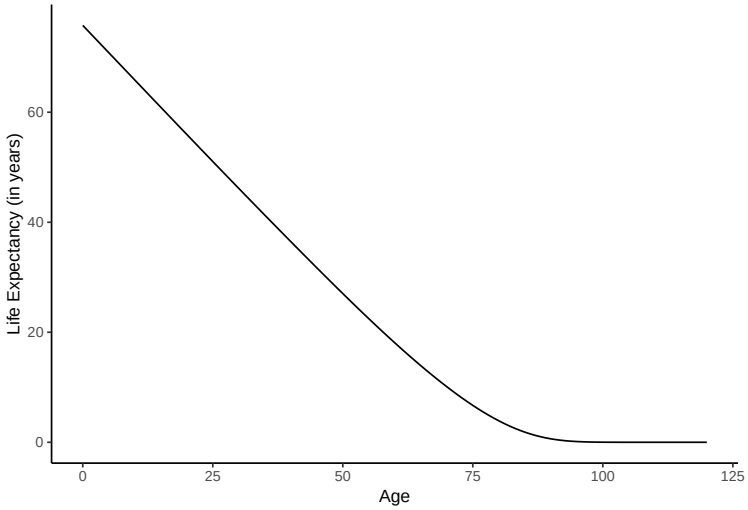


Figure 18: Theoretical Life Expectancy calculated from CSO mortality rates 2015-2017.

Table 8 shows a comparison between the number of deaths in each age bracket from the CSO 2015-2017 life table and the number of deaths that were simulated. The data is premised on a radix of 10,000 for each. This involved dividing the figures from the CSO 2015-17 life table by 10, as the radix used therein was 100,000.

10 Incoporating Time as a Variable into Mortality Models

The heterogeneous Gompertz model discussed previously provides fitted mortality rates for each age, x, but takes no account of time. The model does not account for time as an explanatory variable and were it to be used for projecting or forecasting mortality rates, assumes implicitly, that mortality rates for each cohort is constant over time.

Table 8: Comparison of number of deaths from CSO 2015-17 life table and number of simulated deaths with radix of 10,000.

Age	Deaths	SimDeaths
0-10	37	88
10-20	10	60
20-30	18	118
30-40	39	171
40-50	75	308
50-60	223	724
60-70	478	1332
70-80	1269	2432
80-90	2825	3312
90-100	4148	1397
100+	878	58

Keyfitz and Caswell [12] makes the point that "any forecasting method, whether naive or sophisticated, depends on some function being approximately constant... [including] to suppose that the crude rates will be constant for the next few years."

10.1 Is it reasonable to assume that mortality rates, and thus model parameters, remain constant over time?

As seen above, we fitted parameters to the three and four sub-popualtion heterogeneous Gompertz models using the CSO life tables mortality data for 2015-2017. The four sub-population heterogeneous model was selected as providing the better fit than the three sub-population model. An insurer might legitimately next ask whether that model with those parameters could be utilised in some way for other time periods than the 2015-2017 time period it had been fitted on, for example, as a predictive tool for making predictions on mortality, likely age to death, *etc.* for future time periods.

To test such a hypothesis, actual data from time periods other than 2015-2017 is compared to the hypothetical rates obtained from this model and goodness-of-fit criteria employed to test the hypothesis, *i.e.* whether the model fitted data from other time periods reasonably well. Such goodness-of-fit criteria could include, for example, R^2 or the BIC employed when selecting how many sub-populations to include in the model above.

Irish mortality rates for the years between 1955 and 2015, at 10 year intervals was obtained from the HMD. Individual plots were then created of these mortality rates with a fitted line from the aforementioned model overlaid to see how well the

Table 9: Table of BIC for how well the model fits the Irish mortality data from 1955 to 2015. Source: HMD

1955	1965	1975	1985	1995	2005	2015
58.391	12.54445	-7.368997	-56.06708	-114.9072	-190.1408	-164.1162

hypothetical rates created by the model fitted the actual data from those years. They are shown at Figure 19.

As can be seen, the fitted line from the model fits the actual data better as time progresses from 1955 to 2015, the year from which the data generating the model was obtained. It is also evident that the hypothetical mortality rate is lower than the actual mortality rates for years earlier than 2015. This is intuitively expected, given the increase in life expectancy over the last 60 years. This is sometimes referred to as time selection of mortality rates and begs the question what can be done to account for such, if, as appears, the model parameters, as fitted on one time period, do not remain valid as calendar time varies.

Figure 20 shows the residuals between the data from 1955 to 2015 and the fitted data from the model. It shows that there is a slight positive bias for the years prior to 2005 and, a slight downward bias for the year 2015.

Looking at the Bayesian Information Crieria (BIC) for how well the model fits the data from 1955 to 2015, which is set out at Table 9, it is clear that this corroborates the evidence from the graphs at Figure 19 and Figure 20; as the value of the BIC goes down, the fit improves.

10.2 Time heterogeneity

Heterogeneity in respect of mortality rates is sometimes referred to as selection of mortality rates. In the more general meaning of the term, selection refers to the division of any characteristic of interest of a population into different sub-groups of that population. A type of selection known as time selection takes place when a particular condition varies with calender time. Thus, it is calendar time itself which is inducing the heterogeneity. The most obvious example for life insurance is mortality rates improving with time. An insurance company may try and overcome such by revising its schedule of mortality rates quite frequently or by using a mortality trend assumption to model the improvement in such, over successive time periods.

Keyfitz and Caswell [12] states that "to project mortality... we need to determine not only overall trends but also what is going to happen to the several ages". One method he cited was the relational method developed by Brass [4]. The relational method "embraces changes through time in a flexible description involving two or at most three parameters... Brass found that to go from one life table to another was easy

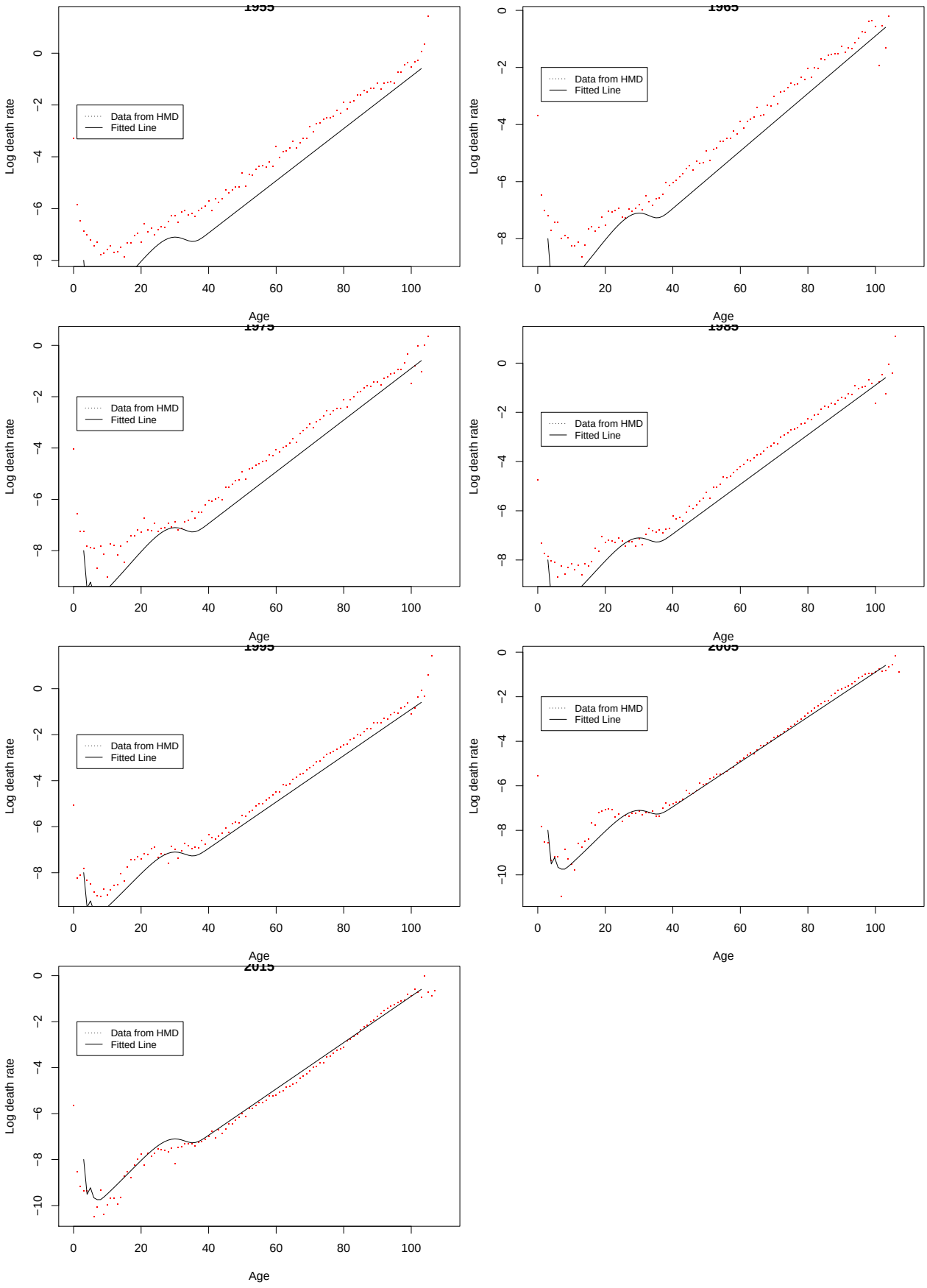


Figure 19: Plots of Irish mortality rates for the years 1955 to 2015 with fitted line from model generated from 2015-2017 CSO data. Source: HMD.

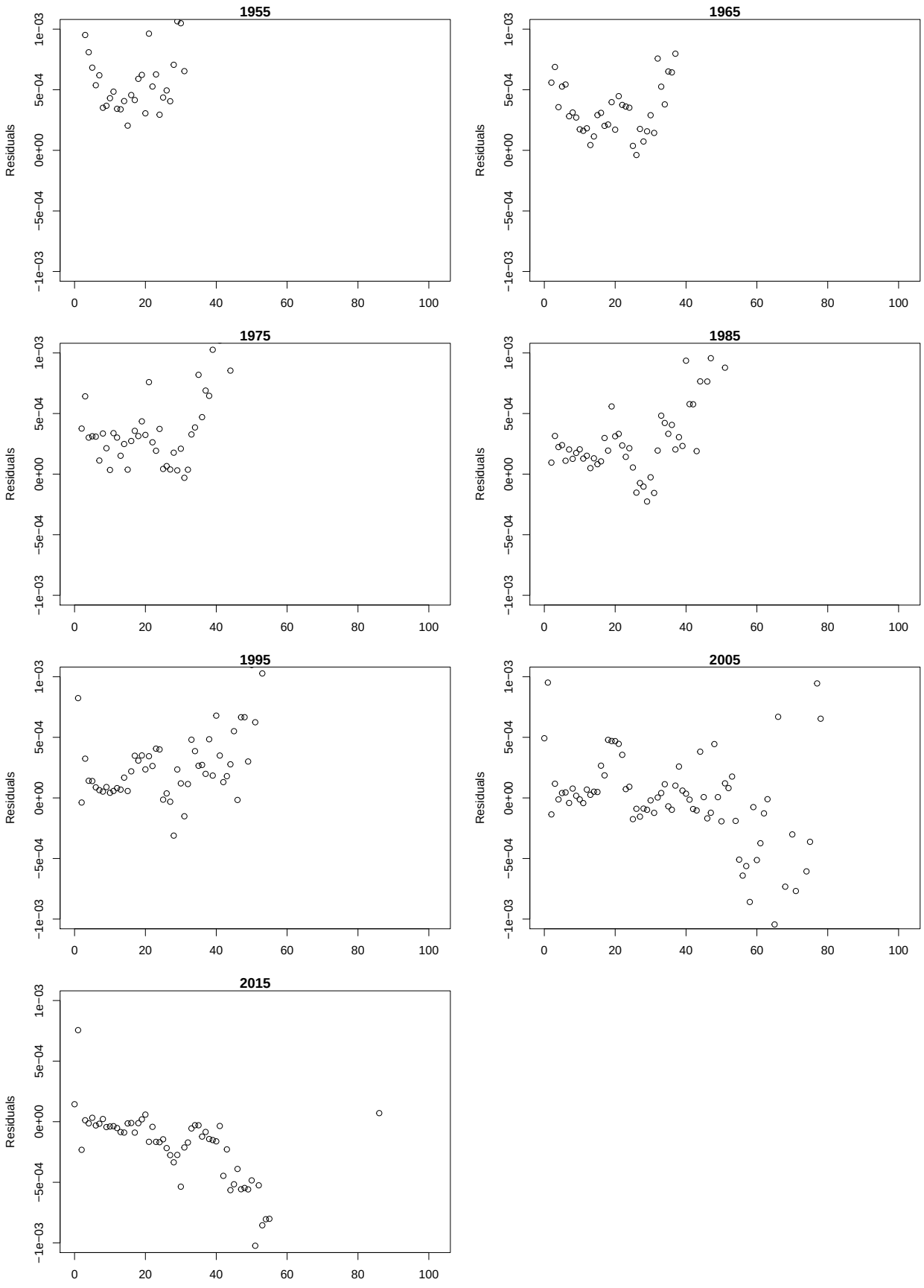


Figure 20: Residuals of data from 1955 to 2015 and fitted data.

if both were transformed into logits, that is their l_x columns were transformed by":

$$Y(x) = \frac{1}{2}\ln[\frac{1-l(x)}{l(x)}] \ . \tag{57}$$

If the early table is distinguished by a subscript s, the relation is the linear

$$Y(x) = \alpha + \beta Y_s(x) \quad \equiv \tag{58}$$

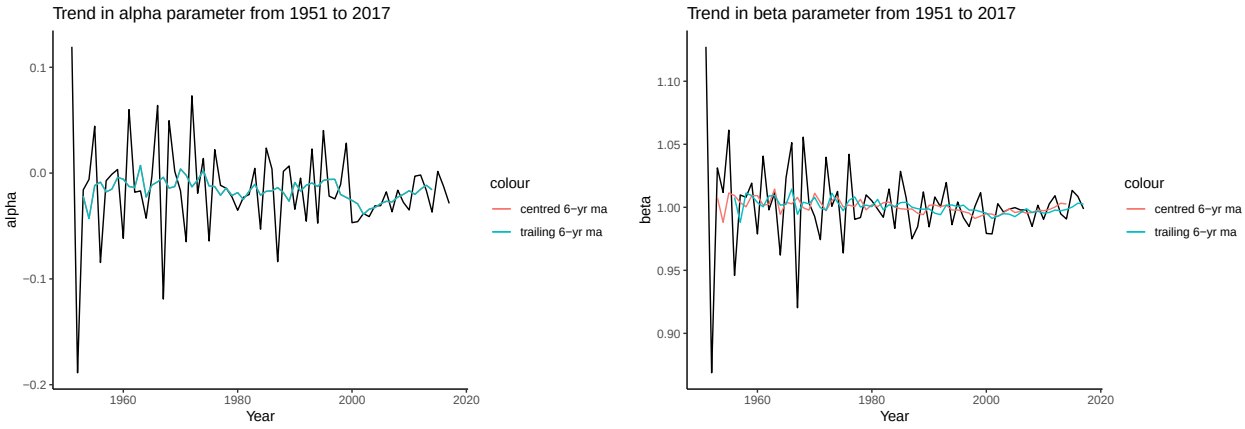
$$\frac{1}{2}\ln[\frac{1-l(x)}{l(x)}] = \alpha + \frac{\beta}{2}\ln[\frac{1-l_s(x)}{l_s(x)}] \tag{59}$$

In forecasting, one would take a life table for time t_s as $l_s(x)$ and fit to it the table for later time t_1, finding the constants α_1 and β_1. A new α_2 and β_2 would be found on fitting the table for t_2 to t_1, and α_3 and β_3 on fitting the table for t_3 to t_2. The several fittings would indicate a trend in α and in β that could then be extrapolated into the future. The extrapolated α and β for future dates, along with the base life table $l_s(x)$, would carry the life table to the dates in question.

This approach was implemented on two datasets:

- Life tables of total mortality from the HMD for Ireland for the consecutive years 1950 to 2017;
- Life tables of total mortality provided by Irish Life Insurance Co. in respect of its policyholders for the years between 1935 and 2015, at 10 year intervals.

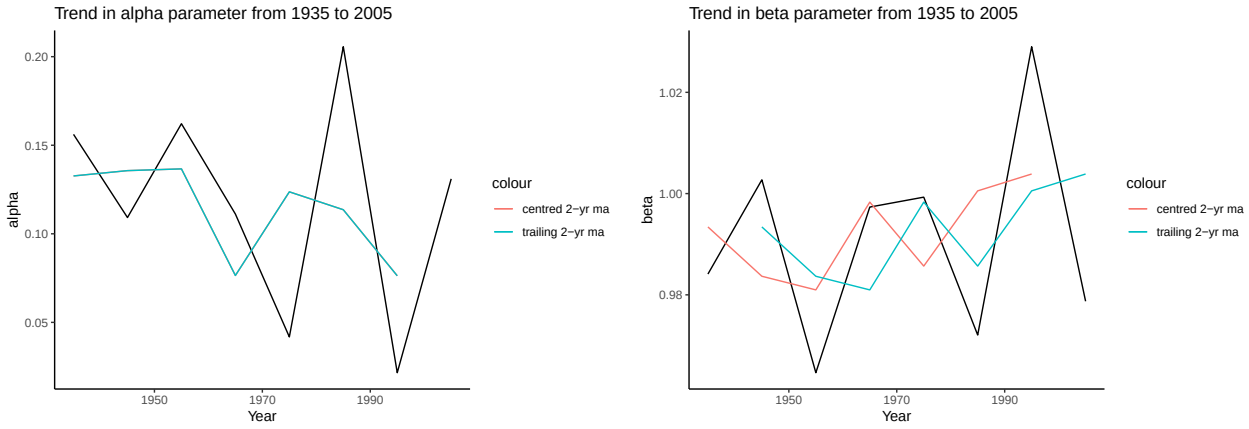
Graphs of the evolution of the alpha and beta parameters over time, together with the moving averages of those parameters, are shown at Figure 21 and Figure 22 for the HMD and Irish Life Insurance Co. datasets respectively.



(a) alpha parameter

(b) beta parameter

Figure 21: Trend in parameters on HMD Irish data



(a) alpha parameter

(b) beta parameter

Figure 22: Trend in parameters on Irish Life Co. data

The mean of the α parameter over the time period 1950 to 2017 for the HMD Irish mortality data is -0.0159 and the mean of the β parameter for the same dataset is 1.0013. The mean of the α parameter over the time period 1935 to 2005 for the Irish Life Insurance Co. mortality data is 0.1173 and the mean of the β parameter for the same dataset is 0.991.

Studying mortality trends of the past half century in England and Wales, Brass considers it possible that in the future β may well come close to unity, and α decrease by 0.5 every 40 years. The data from the two Irish life tables appears to provide a close fit to this prediction in respect of the British data, at least for the β parameter.

10.2.1 Incorporating the α and β parameters into the heterogeneous Gompertz model

One way of incorporating a time dimension into the heterogeneous Gompertz model is to make an allowance for the α and β parameters into the heterogeneous Gompertz model. This would be a fruitful topic for further research.

11 Conclusion

The heterogeneous Gompertz model fits life table data from standard public access sources such as the CSO and HMD quite well. The model appears to be overly complicated for mortality data provided by the Irish Life insurance company, which appears to be a far more homogeneous cohort than the broader population from which it is drawn. A simple, one population, homogeneous Gompertz model would appear to fit that data adequately.

The fitted model appears to be limited in its temporal validity however, as the model does not appear to fit mortality data from older periods as well as the period data from which it's parameters are fitted to. Whether that feature is true into the future is an open question, for the obvious reason that the data is not available on which to test the model as yet. One way that question could be approached is seeing whether there is a relatively straightforward trend to the variation of the parameters, such as linear growth. The answer to that question appears to be in the positive. Whether that can be extrapolated into the future is, again, an open question, given some of the slowdowns and even reversals in the growth of life expectancy witnessed in some developed countries in recent years. It is too early to say whether a ceiling has been reached in relation to the growth in life expectancy, and hence the mortlaity rate, μ_x, in developed countries but this could be an avenue for further research.

12 References

[1] M. E. Atkinson and D. C. M. Dickson. *An introduction to actuarial studies*. Cheltenham, UK ; Northampton, MA: Edward Elgar Pub, 2000. ISBN: 978-1-84064-446-3.

[2] Demetris Avraam, Joao Pedro de Magalhaes, and Bakhtier Vasiev. "A mathematical model of mortality dynamics across the lifespan combining heterogeneity and stochastic effects". In: *Experimental Gerontology* 48.8 (Aug. 2013), pp. 801–811. ISSN: 0531-5565. DOI: 10.1016/j.exger.2013.05.054. URL: https://www.sciencedirect.com/science/article/pii/S0531556513001885.

[3] Demetris Avraam et al. "On the heterogeneity of human populations as reflected by mortality dynamics". In: *Aging* 8 (Nov. 2016). DOI: 10.18632/aging.101112.

[4] W. Brass. "Perspectives in Population Prediction: Illustrated by the Statistics of England and Wales". In: *Journal of the Royal Statistical Society. Series A (General)* 137.4 (1974). Publisher: [Royal Statistical Society, Wiley], pp. 532–583. ISSN: 00359238. DOI: 10.2307/2344713. URL: http://www.jstor.org/stable/2344713 (visited on 04/06/2022).

[5] James R. Carey and Deborah Ann Roach. *Biodemography: an introduction to concepts and methods*. OCLC: on1126117223. Princeton, New Jersey: Princeton University Press, 2020. ISBN: 978-0-691-12900-6.

[6] Mike Christie, ed. *Simplicity, complexity, and modelling*. Statistics in practice. OCLC: ocn739645795. Chichester, West Sussex: Wiley, 2011. ISBN: 978-0-470-74002-6.

[7] Regina C. Elandt-Johnson and Norman Lloyd Johnson. *Survival models and data analysis*. Wiley series in probability and mathematical statistics. New York: Wiley, 1980. ISBN: 978-0-471-03174-1.

[8] *Evolutionary and swarm intelligence algorithms*. New York, NY: Springer Berlin Heidelberg, 2018. ISBN: 978-3-319-91339-1.

[9] B. Gompertz. "On the nature of the function expressive of the law of human mortality and on a new model of determining life contingencies." In: *Phil. Trans. R. Soc.* 115 (1825), pp. 513–585.

[10] Jessica Y Ho and Arun S Hendi. “Recent trends in life expectancy across high income countries: retrospective observational study”. In: *BMJ* 362 (2018). Publisher: BMJ Publishing Group Ltd _eprint: https://www.bmj.com/content/362/bmj.k2562.full.pdf. ISSN: 0959-8138. DOI: 10.1136/bmj.k2562. URL: https://www.bmj.com/content/362/bmj.k2562.

[11] *Human Mortality Database*. URL: www.mortality.org.

[12] Nathan Keyfitz and Hal Caswell. *Applied mathematical demography*. 3rd ed. Statistics for biology and health. New York, NY: Springer, 2005. ISBN: 978-0-387-22537-1.

[13] Thomas B L Kirkwood. “Deciphering death: a commentary on Gompertz (1825) ’On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies’”. eng. In: *Philos Trans R Soc Lond B Biol Sci* 370.1666 (Apr. 2015). Publisher: The Royal Society, p. 20140379. ISSN: 1471-2970. DOI: 10.1098/rstb.2014.0379. URL: https://pubmed.ncbi.nlm.nih.gov/25750242.

[14] Ermanno Pitacco. “Heterogeneity in mortality: a survey with an actuarial focus”. In: *European Actuarial Journal* 9.1 (July 2019), pp. 3–30. ISSN: 2190-9741. DOI: 10.1007/s13385-019-00207-z. URL: https://doi.org/10.1007/s13385-019-00207-z.

[15] Ermanno Pitacco, ed. *Modelling longevity dynamics for pensions and annuity business*. OCLC: ocn243820394. Oxford ; New York: Oxford University Press, 2009. ISBN: 978-0-19-954727-2.

[16] Pletcher. “Model fitting and hypothesis testing for age-specific mortality data”. In: *Journal of Evolutionary Biology* 12.3 (May 1999). Publisher: John Wiley & Sons, Ltd, pp. 430–439. ISSN: 1010-061X. DOI: 10.1046/j.1420-9101.1999.00058.x. URL: https://doi.org/10.1046/j.1420-9101.1999.00058.x (visited on 04/04/2022).

[17] J. H. Pollard. *Mathematical Models for the Growth of Human Populations*. Cambridge: Cambridge University Press, 1978. ISBN: 978-0-521-29442-3.

[18] Nathan Rooy. *Simple Differential Evolution with Python*. URL: https://nathanrooy.github.io/posts/2017-08-27/simple-differential-evolution-with-python/.

[19] Gideon Schwarz. “Estimating the Dimension of a Model”. In: *The Annals of Statistics* 6.2 (1978). Publisher: Institute of Mathematical Statistics, pp. 461–464. ISSN: 00905364. URL: http://www.jstor.org/stable/2958889 (visited on 04/04/2022).

[20] Mara Tableman and Jong Sung Kim. *Survival analysis using S: analysis of time-to-event data*. Texts in statistical science. Boca Raton, Fla: Chapman & Hall/CRC, 2004. ISBN: 978-1-58488-408-8.

[21] James W. Vaupel, Kenneth G. Manton, and Eric Stallard. “The impact of heterogeneity in individual frailty on the dynamics of mortality”. In: *Demography* 16.3 (Aug. 1979), pp. 439–454. ISSN: 0070-3370. DOI: 10.2307/2061224. URL: https://doi.org/10.2307/2061224 (visited on 02/18/2022).

[22] James W. Vaupel and Trifon I. Missov. “Unobserved population heterogeneity: A review of formal relationships”. In: *Demographic Research* 31.22 (Sept. 2014), pp. 659–686. ISSN: 1435-9871. URL: https://www.demographic-research.org/volumes/vol31/22/.

[23] Anatoli I. Yashin, Ivan A. Iachine, and Alexander S. Begun. “Mortality modeling: A review”. In: *null* 8.4 (Oct. 2000). Publisher: Routledge, pp. 305–332. ISSN: 0889-8480. DOI: 10.1080/08898480009525489. URL: https://doi.org/10.1080/08898480009525489.

[24] Yudong Zhang, Shuihua Wang, and Genlin Ji. “A Comprehensive Survey on Particle Swarm Optimization Algorithm and Its Applications”. In: *Mathematical Problems in Engineering* 2015 (Oct. 2015). Ed. by Shuming Wang. Publisher: Hindawi Publishing Corporation, p. 931256. ISSN: 1024-123X. DOI: 10.1155/2015/931256. URL: https://doi.org/10.1155/2015/931256.

13 Appendix: All code for this thesis

The R code used in this thesis is printed below. In addition, a zip file will be uploaded containing all of the R Markdown files (.Rmd) used to create the final report, including all sub-directories containing images, custom packages, data, *etc.* Only the `Master.Rmd` file is ‘knitted’ by RStudio. This file essentially collates all the other .Rmd files into one pdf (or html) output.

```
knitr::opts_chunk$set(echo = FALSE, message=FALSE, warning = FALSE, comment='')
if(!require(tidyverse)) install.packages("tidyverse")
```

```r
if(!require(tufte)) install.packages("tufte")
if(!require(stringr)) install.packages("stringr")
if(!require(embedr)) install.packages("embedr")
if(!require(demography)) install.packages("demography")
if(!require(DiagrammeR)) install.packages("DiagrammeR")
if(!require(cowplot)) install.packages("cowplot")
if(!require(devtools)) install.packages("devtools")
if(!require(stocks)) install.packages("stocks")
if(!require(csodata)) install.packages("csodata")
if(!require(lemon)) install.packages("lemon")
if (knitr::is_html_output()) {
    library(embedr)
  # embed .mp4 video
  video <- "examples/animation.mp4"
  embed_video(video, type="mp4", width = "400", height = "350")
} else {
  # # download Irish mortality rates from mortality.org HMD
  library(demography)
  irl <- hmd.mx("IRL", "paul.christopher@mycit.ie", "R00207143", "Ireland")
  # implementation of 3d plot from Charpentier, p.324 for Ireland - observed only
  irl1950 <- extract.years(irl, years=1950:2017)
  year = irl1950$year[seq(1,65,by=3)]
  age = irl1950$age[seq(1,100,by=3)]
  persp(age, year, log(irl1950$rate$male[seq(1,100, by=3), seq(1,65,by=3)]),
        theta = -30, main='Observed death rates', col=grey(.93), shade=TRUE,
        xlab = "Age", ylab = "", zlab = "Log death rate", ticktype = "detailed",
        cex.axis=0.6, cex.lab=0.6)
}
#create life expectancy dataframe
e_0_irl_all = data.frame(Year=integer(0), Male=numeric(0), Female=numeric(0), Total=numer
for (year in seq(1950,2015)){
  output = c(year, lifetable(irl, years = year, series = "male")$ex[1],
             lifetable(irl, years = year, series = "female")$ex[1],
             lifetable(irl, years = year, series = "total")$ex[1])
  names(output) <- names(e_0_irl_all)
  e_0_irl_all = rbind(e_0_irl_all, as.list(output))
}
```

```r
# plot life expectancy
library(lemon)
ggplot(e_0_irl_all, aes(x=Year)) +
  geom_line(aes(y = Total,  linetype='total')) +
  geom_line(aes(y = Male, linetype='male')) +
  geom_line(aes(y = Female, linetype='female')) +
  theme_classic() +
  theme(legend.title=element_blank()) +
  ylab("Life Expectancy at birth") +
  lemon::coord_capped_cart( left = 'both')
library(DiagrammeR)
DiagrammeR::grViz("
  digraph {
  layout = dot
  node [shape = circle]
  0 -> 1 [taillabel = 'Survive']
  die1 [label='Die', shape=none]
  0 -> die1
  1 -> 2 [taillabel = 'Survive']
  die [label= 'Die', shape=none]
  1 -> die
}")
knitr::include_graphics(c("images/CSO_LT_1.png", "images/CSO_LT_2.png"))
library(demography)
irl <- hmd.mx("IRL", "paul.christopher@mycit.ie", "R00207143", "Ireland")
plot(lifetable(irl, years = 2013, series = "female")$lx, type='l', xlab = 'Age', ylab='S(
lines(lifetable(irl, years = 2013, series = "male")$lx, type='l', lty=2)
lines(lifetable(irl, years = 2013, series = "total")$lx, type='l', lty=2, lwd=3)
legend(1, 0.5, legend=c("Female", "Total", "Male"), lty=c(1,2,2), lwd=c(1,3,1), cex=0.8)
library(demography)
library(tidyverse)
library(stocks)
# download Irish mortality rates from mortality.org HMD
irl <- hmd.mx("IRL", "paul.christopher@mycit.ie", "R00207143", "Ireland")
irl_15 <- extract.years(irl, 2015)
irl_mx <- irl_15$rate$total[1:100] # create vector from Irish data
irl_qx <- (1-exp(-irl_mx)) # create vector of approx qx from mx
Age <- 0:99 # generate a vector of ages from 0 to omega
```

```
dat <- as.data.frame(cbind(Age=Age, irl_qx=irl_qx)) # bind the vectors into dataframe
dat$Ratio <- c(ratios(irl_qx), ratios(irl_qx)[99])
# Plots
p1 <- ggplot(dat, aes(x=Age, y=irl_qx)) + geom_line() +theme_classic() +
  ylab("Mortality (q(x))")
p2 <- ggplot(dat, aes(x=Age, y=Ratio)) + geom_line() + ylab("Ratio q(x+1)/q(x)") +
  theme_classic()
p1 + annotation_custom(ggplotGrob(p2), xmin = 10, xmax=75, ymin = 0.06, ymax=0.27)
library(kableExtra)
kable(dat[71:76,], caption="Ireland (2015) Mortality, q(x) and Rate of Change \\label{rat
knitr::include_graphics('./images/risk.png')
library(devtools)
install("projectRpkg")
library(tidyverse)
age_vec <- 0:100 # create vector of ages from 0 to 100

Avr_A <- read_csv("data/Avraam_graphA.csv")
# Plot B - Avraam
rho0 <- c(0.5, 0.5) # sub-pop proportions
m0 <- c(0.03, 0.03) # initial mort rates for each sub-pop
beta <- c(0.067, 0.033) # ageing rates for each sub-pop
# create vector of total population mortality rates
ma <- projectRpkg::gompertz(rho0, m0, beta, age_vec)
mb <- projectRpkg::gompertz(rho0, m0, beta, age_vec)
mc <- projectRpkg::gompertz(rho0, m0, beta, age_vec)
# create matrix of sub-population mortality rates for sub-pops 2, 1a, 1b, 1c
m1a <- 0.03*exp(0.2*age_vec) # sub-pop1 = blue
m2 <- 0.03*exp(0.033*age_vec) # sub-pop2
m1b <- 0.03*exp(0.1*age_vec) # sub-pop1 = green
m1c <- 0.03*exp(0.067*age_vec) # sub-pop1 = red
Avr_B <- as.tibble(cbind("Age"=age_vec, ma, mb, mc, m1a, m1b, m1c, m2))
# Plot C - Avraam
rho0a <- c(0.9, 0.1)
rho0b <- c(0.99, 0.01)
rho0c <- c(0.999, 0.001) # sub-pop proportions
m0 <- c(0.15, 0.02) # initial mort rates for each sub-pop
beta <- c(0.036, 0.056) # ageing rates for each sub-pop
# create vector of total population mortality rates
```

```
ma <- projectRpkg::gompertz(rho0a, m0, beta, age_vec)
mb <- projectRpkg::gompertz(rho0b, m0, beta, age_vec)
mc <- projectRpkg::gompertz(rho0c, m0, beta, age_vec)
# create matrix of sub-population mortality rates for sub-pops 2, 1a, 1b, 1c
m1 <- 0.15*exp(0.036*age_vec) # sub-pop1 = blue
m2 <- 0.02*exp(0.056*age_vec) # sub-pop2
Avr_C <- as.tibble(cbind("Age"=age_vec, ma, mb, mc, m1, m2))
plotA <- ggplot(Avr_A, aes(x=Age)) +
  xlim(c(0,60)) + ylim(c(0.00001,1)) +
  geom_line(aes(y = m_a, colour = "m1=0.5")) +
  geom_line(aes(y = m_b, colour = "m1=0.25")) +
  geom_line(aes(y=m_c, colour = "m1=0.15")) +
  geom_line(aes(y=m2, colour = "m2=0.02"), linetype='dashed', colour='black') +
  geom_line(aes(y=m1a, colour="m1=0.5"), linetype='dashed') +
  geom_line(aes(y=m1b, colour="m1=0.25"), linetype='dashed') +
  geom_line(aes(y=m1c, colour="m1=0.15"), linetype='dashed') +
  scale_y_continuous(trans = 'log', breaks = c(.02, .1,1)) + ylab("Mortality Rate") +
  theme_classic()

plotB <- ggplot(Avr_B, aes(x=Age)) +
  xlim(c(0,60)) + ylim(c(0.01,1)) +
  geom_line(aes(y = ma, colour = "beta=0.2")) +
  geom_line(aes(y = mb, colour = "beta=0.1")) +
  geom_line(aes(y=mc, colour = "beta=0.067")) +
  geom_line(aes(y=m2, colour = "m2=0.02"), linetype='dashed', colour='black') +
  geom_line(aes(y=m1a, colour= "beta=0.2"), linetype='dashed') +
  geom_line(aes(y=m1b, colour="beta=0.1"), linetype='dashed') +
  geom_line(aes(y=m1c, colour="beta=0.067"), linetype='dashed') +
  scale_y_continuous(trans = 'log', breaks = c(0.03, .1,1)) + ylab("Mortality Rate") +
  theme_classic()

plotC <- ggplot(Avr_C, aes(x=Age)) +
  xlim(c(0,60)) + ylim(c(0.001,1)) +
  geom_line(aes(y = ma, colour = "rho=0.9")) +
  geom_line(aes(y = mb, colour = "rho=0.99")) +
  geom_line(aes(y=mc, colour = "rho=0.999")) +
  geom_line(aes(y=m2, colour = "rho2=0.1"), linetype='dashed', colour='black') +
  geom_line(aes(y=m1, colour= "rho=0.9"), linetype='dashed', colour='black') +
```

```
  scale_y_continuous(trans = 'log', breaks=c(.05,.1,1)) + ylab("Mortality Rate") +
  theme_classic()
library(cowplot)
plot_grid(plotA, plotB, plotC, labels=c("A","B","C"), ncol = 1, nrow = 3)
library(demography)
# download 2007 Swedish mortality rates from mortality.org HMD
sweden <- hmd.mx("SWE", "paul.christopher@mycit.ie", "R00207143", "Sweden")
# create heterogeneous gompertz model
age_vec <- 0:110
rho0 <- c(0.00198, 0.00483, 0.99319) # 3 sub-pop proportions
m0 <- c(0.7211, 0.001169, 0.00001317) # initial mort rates for each sub-pop
beta <- c(.67e-5, 0.2129, 0.1041) # ageing rates for each sub-pop
swe_mrates <- projectRpkg::gompertz(rho0,m0, beta, age_vec)

#plot the data and fitted model on same plot
plot(age_vec, log(swe_mrates), type='l', xlab='Age', ylab="Log of mortality rates") # plo
points(sweden, year=2007, series='total', col='red', cex=.9) # overlay the actual data
swe07_params <- readxl::read_excel("data/swe07_params.xls", col_names = FALSE)
colnames(swe07_params) <- c("Parameters", "Values", "Avraams's values")
knitr::kable(swe07_params,caption = "Parameters for 4 sub-population heterogeneous Gomper
library(tidyverse)

age_vec <- 0:102 # create vector of ages from 0 to 100
# Swedish 2007 data from HMD
swe <- read.table("data/SWE.bltper_1x1.txt", skip=2, header=TRUE)
swe07 <- swe %>% filter(Year == 2007)
swe07 <- swe07[1:103,]
rho0 <- swe07_params$Values[c(3,6,9,12)] # sub-pop proportions
m0 <- swe07_params$Values[c(1,4,7,10)] # initial mort rates for each sub-pop
beta <- swe07_params$Values[c(2,5,8,11)] # ageing rates for each sub-pop
# create vector of total population mortality rates from R model using above parameters
gom_mod <- projectRpkg::gompertz(rho0, m0, beta, age_vec)
# The plot
plot(age_vec, log(swe07$qx), xlab="Age", ylab = "Log of Mortality Rate") # plot actual Sw
lines(log(gom_mod)) # add fitted line from R Gompertz model
legend(1,-2, legend = c("Data from HMD", "Fitted Line"), lty = c(3,1), cex=0.8)

# Same for Irish data
```

```r
irl <- readxl::read_xls("data/ILT15.xls") # the actual data from CSO
length(irl$q_x[1:102])

rho0_4 <- c(0.002547208824796, 0.000650784396633, 0.006435402652686, 0.990366604125885) #
m0_4 <- c(3.20393321656011, 0.282271320800018, 0.000570813106869, 1.717315579000E-05) # i
beta_4 <- c(0.010226506161476, 0.221852537463537, 0.206653841941924, 0.100761335928826) #
# create vector of total population fitted mortality rates
irl_tot_4 <- projectRpkg::gompertz(rho0_4, m0_4, beta_4, age_vec)
# The plot
plot(age_vec, log(irl$q_x), xlab="Age", ylab = "Log of Mortality Rate") # plot actual dat
lines(log(irl_tot_4)) # add fitted line from R gompertz model
legend(1, -2, legend=c("Data from CSO", "Fitted Line"), lty=c(3,1), cex=0.8)

library(devtools)
install("projectRpkg")
age_vec <- 0:102 # create vector of ages from 0 to 102
y <- readxl::read_xls("data/ILT15.xls") # the actual data

# Fitted 3 pop model
rho0 <- c(0.002237328077671, 0.000557174246818, 0.997205497675511) # sub-pop proportions
m0 <- c(4.54973908257018, 0.368200387234967, 3.568107751643E-05) # initial mort rates for
beta <- c(-0.1500761318802, 0.211537687783045, 0.09139640495285) # ageing rates for each
# create vector of total population mortality rates
m_tot <- projectRpkg::gompertz(rho0, m0, beta, age_vec)
# create matrix of sub-population mortality rates for sub-pops 1,2 3
m1 <- 4.54973908257018*exp(-0.15007613188022*age_vec[0:30]) # sub-pop1
m2 <- 0.368200387234967*exp(0.211537687783045*age_vec[0:30]) # sub-pop2
m3 <- 3.56810775164346E-05*exp(0.09139640495285*age_vec[0:30]) # sub-pop3
m1 <- append(m1, rep(NA,73), after = length(m1))
m2 <- append(m2, rep(NA,73), after = length(m2))
m3 <- append(m3, rep(NA,73), after = length(m3))
Irl_fit <- as.tibble(cbind("Age"=age_vec, "Data"=y, m_tot, m1, m2, m3))
# Fitted 4 pop model
rho0_4 <- c(0.002547208824796, 0.000650784396633, 0.006435402652686, 0.990366604125885) #
m0_4 <- c(3.20393321656011, 0.282271320800018, 0.000570813106869, 1.717315579000E-05) # i
beta_4 <- c(0.010226506161476, 0.221852537463537, 0.206653841941924, 0.100761335928826) #
# create vector of total population mortality rates
m_tot_4 <- projectRpkg::gompertz(rho0, m0, beta, age_vec)
```

```
# create matrix of sub-population mortality rates for sub-pops 2, 1a, 1b, 1c
m1_4 <- 3.20393321656011*exp(0.010226506161476*age_vec[0:30]) # sub-pop1
m2_4 <- 0.282271320800018*exp(0.221852537463537*age_vec[0:30]) # sub-pop2
m3_4 <- 0.000570813106869*exp(0.206653841941924*age_vec[0:30]) # sub-pop3
m4_4 <- 1.71731557900086E-05*exp(0.100761335928826*age_vec[0:30]) # sub-pop3
m1_4 <- append(m1_4, rep(NA,73), after = length(m1_4))
m2_4 <- append(m2_4, rep(NA,73), after = length(m2_4))
m3_4 <- append(m3_4, rep(NA,73), after = length(m3_4))
m4_4 <- append(m4_4, rep(NA,73), after = length(m4_4))
Irl_fit_4 <- as.tibble(cbind("Age"=age_vec, "Data"=y, m_tot_4, m1_4, m2_4, m3_4, m4_4))

params_rho <- c("$\\rho_{01}$", "$\\rho_{02}$", "$\\rho_{03}$")
params_m0 <- c("$m_{01}$", "$m_{02}$","$m_{03}$")
params_beta <- c("$\\beta_{1}$", "$\\beta_{2}$","$\\beta_{3}$")
param_df <- as.data.frame(cbind(params_rho, rho0, params_m0, m0, params_beta, beta))

kable_table <- knitr::kable(param_df, col.names = c("$\\rho_{0i}$","" ,"$m_{0i}$","" , "$

kable_table

params_rho4 <- c("$\\rho_{01}$", "$\\rho_{02}$", "$\\rho_{03}$", "$\\rho_{04}$")
params_m04 <- c("$m_{01}$", "$m_{02}$","$m_{03}$","$m_{04}$")
params_beta4 <- c("$\\beta_{1}$", "$\\beta_{2}$","$\\beta_{3}$","$\\beta_{4}$")
param_df4 <- as.data.frame(cbind(params_rho4, rho0_4, params_m04, m0_4, params_beta4, beta

kable_table <- knitr::kable(param_df4, col.names = c("$\\rho_{0i}$","" ,"$m_{0i}$","" , "$
styled_kable_table <- kableExtra::kable_styling(kable_table)

#kable_table
styled_kable_table
knitr::include_graphics("images/CSO_fit_3pop.png")
knitr::include_graphics("images/CSO_fit_4pop.png")

knitr::include_graphics(c("images/IrishLifeCo_fit_3pop.png","images/IrishLifeCo_fit_4pop.p

data=read.csv("data/q_Ireland_by_Years.csv")
# create new columns containing the log-transformed mortality rates
data %>% mutate_at(vars(matches("X[0-9]{4}")), list(ln=log)) -> data2
```

```r
colnames(data2)[1] <- 'Age'

m1=lm(data2$X2015_ln~data2$Age)
summary(m1)
# Select columns from data2 if their names start with "expression"
loop.vector <- names(data2[, grep(pattern="X[0-9]{4}_ln", colnames(data2))])
lm.test <- vector("list", length(loop.vector)) #create empty list to store models
for(i in seq_along(loop.vector)){
  lm.test[[i]] <- lm(reformulate("Age", loop.vector[i]), data = data2)
}

cfs <- lapply(lm.test, coef) # list of model coefficients
names(cfs) <- loop.vector # give them names
cfs <- t(as.data.frame(cfs))
rownames(cfs) <- c('2015','2005','1995','1985','1975','1965','1955','1945','1935')
knitr::kable(cfs, caption = "Evolution of parameters of Gompertz model for Irish Life ins
# Use these coefficients to fit lines to data for the years
# create vector/matrix of fitted values to plot with
data1 <- list()

for(i in rownames(cfs)){
  data1[[paste("fit", i, sep="")]] <- cfs[i,1]+cfs[i,2]*data2$Age
}

data2 <- cbind(data2, data1)

# plot the fitted lines for 1935-2015 on one plot
# first re-shape the data to long form
data3 <- data2 %>% pivot_longer(cols= starts_with("fit"), names_to = "year", values_to="q

ggplot(data3, aes(Age,q, color=year)) + geom_line() + theme_classic() +
  labs(y = "Log of Mortality Rate")
age_vec <- 0:110 # create vector of Ages: 0-110
# create empty vector to store life expectancy at birth
life_exp <- vector("list", length(lm.test))
for(i in seq_along(lm.test)){
  ln_q_hat <- predict(lm.test[[i]], Age = age_vec)
  p1 <- 1
```

```
  p_hat <- Reduce(function(v, x) v*(1-x), x=exp(ln_q_hat), init=p1, accumulate=TRUE)
  life_exp[[i]] <- sum(p_hat[2:length(p_hat)])
}
names(life_exp) <- loop.vector # add names to elements of the vector
th_life_exp_df <- as.data.frame(life_exp)
colnames(th_life_exp_df) <- c('2015','2005','1995','1985','1975','1965','1955','1945','19
knitr::kable(round(rev(th_life_exp_df),2), caption = "Theoretical life expectancy calcula
library(demography)
irl <- hmd.mx("IRL", "paul.christopher@mycit.ie", "R00207143", "Ireland")
irl.lt <- lifetable(irl, series = names(irl$rate[3]))
df1 <- as.data.frame(irl.lt$Tx[1, seq(6,68, by=10)])
colnames(df1) = c('')
knitr::kable(round(t(df1),2), caption = "Actual life expectancy in Ireland between 1955 a
# library(csodata)
# tbl1 <- cso_get_data('VSA30')
library(lemon)
tbl1 <- readRDS("data/VSA30.Rda")
tbl1[1:2,] %>% pivot_longer(cols=matches("[0-9]{4}"), names_to="Year", values_to="count")
tbl2$Year <- as.numeric(tbl2$Year)
ggplot(tbl2, aes(x=Year, y=count, colour=Sex)) + geom_line() + ylab('Life Expectancy at b
  theme_classic() + ylim(c(40,100)) +
  lemon::coord_capped_cart( left = 'both')
plot(irl.lt)
life_ex = readxl::read_xlsx("data/LifeExp.xlsx")

AgeDeath=0
for (i in 1:10000){
  alive=1;t=0; while (alive==1) {
    t=t+1
    if (runif(1,0,1)< life_ex$`q(x)`[t]) {AgeDeath[i]=t-1;alive=0}
  }
}
SimDeaths <- hist(AgeDeath, main = NULL)
ggplot(life_ex, aes(life_ex$`Age (x)`, life_ex$`T(x)`, )) + geom_line() +
  theme_classic() + xlab("Age") + ylab("Life Expectancy (in years)")
# Compare simulated deaths to CSO data from 2015-17
CSOdeaths <- readxl::read_xls("data/NumDeaths15.xls")
# NumDeaths15.xls is extract from CSO 2015-17 life table showing num of deaths pa from a
```

```r
CSOdeaths$`Num deaths` <- CSOdeaths$`Num deaths`/10 # as CSO life table is 100,000 lives
# bin this data into 10 yr intervals and aggregate
CSOdeaths$Age <- cut(CSOdeaths$Age, 11, labels=F)
AgeLabels <- c("0-10",'10-20','20-30','30-40','40-50','50-60','60-70','70-80','80-90','90
NumDeaths <- aggregate(`Num deaths`~Age, data=CSOdeaths, sum)
NumDeaths$Age <- AgeLabels
# extract binned deaths from simulation
# Wrap binned data from sim and 2015 into dataframe
AgeDeath_df <- data.frame(Age=NumDeaths$Age,Deaths=round(NumDeaths$`Num deaths`), SimDeatl
knitr::kable(AgeDeath_df, caption = "Comparison of number of deaths from CSO 2015-17 life
library(demography)
age_vec <- seq(0:102)
rho0 <- c(0.002547208824796, 0.000650784396633, 0.006435402652686, 0.990366604125885) # s
m0 <- c(3.20393321656011, 0.282271320800018, 0.000570813106869, 1.717315579000E-05) # ini
beta <- c(0.010226506161476, 0.221852537463537, 0.206653841941924, 0.100761335928826) # a
# create vector of total population fitted mortality rates
irl_tot_4 <- projectRpkg::gompertz(rho0, m0, beta, age_vec)
# download 2007 Irish mortality rates from mortality.org HMD
irl <- hmd.mx("IRL", "paul.christopher@mycit.ie", "R00207143", "Ireland")
# The plots
#par(mfrow=c(1,2))
par(mar = c(4, 4, .1, .1))
for(i in seq(1955, 2015, 10)){
  plot(irl, year=i, series='total', type='p', pch=".", cex=2, col='red', main=i) # plot i
  lines(log(irl_tot_4))
  legend(1,-2, legend=c("Data from HMD", "Fitted Line"), lty=c(3,1), cex=0.8)
}

# Plot residuals
# set up dataframe to hold residuals
residdf <- setNames(data.frame(matrix(ncol = 7, nrow = 103)), seq(1955,2015,10))
for(i in seq(1955, 2015, 10)){
  residdf[,paste("",i,sep='')] <- (irl$rate$total[1:103,paste("",i,sep='')] - irl_tot_4)
}
residdf$Age <- seq(from=0, to=102)
# The plots
par(mar = c(3.1, 4.1, 1.1, 1.1))
for(i in seq(1955, 2015,10)){
```

```
  plot(residdf$Age, residdf[,paste('',i,sep='')], ylim = c(-0.001, .001),
       ylab = "Residuals", xlab = 'Age', main=i)
}
#BIC
for( i in seq(1955,2015,10)){
  residdf[,paste("RS",i,sep='')] <- (log(irl$rate$total[1:103,paste('',i,sep='')])-log(ir
}
#list of BICs
BICs <- list()
for(i in seq(1955,2015,10)){
  RSS <- sum(residdf[,paste("RS",i,sep='')], na.rm = T)
  name <- paste('BIC:',i,sep='')
  BICs <- append(BICs, 103*log(RSS/103)+11*log(103) )
}

BICs <- data.frame(BICs)
colnames(BICs) <- seq(1955,2015,10)

knitr::kable(BICs, caption = "Table of BIC for how well the model fits the Irish mortalit
#%>% kable_styling(latex_options = "striped")

library(demography)
# transform l_x to logits
irl <- hmd.mx("IRL", "paul.christopher@mycit.ie", "R00207143", "Ireland")
l_x <- lifetable(irl, series = "total")$lx
y_x <- 0.5*log((1-l_x)/l_x)
y_x <- y_x[-1,] # 1st age is -Inf, remove it
# make linear model and gather a and b parameters
coeff_vecHMD <- matrix(, nrow=67, ncol=2)
for (index in 2:length(y_x[1,])){
  coeff_vecHMD[index-1, 1] <- lm(y_x[,index] ~ y_x[,index-1])$coefficients[[1]]
  coeff_vecHMD[index-1, 2] <- lm(y_x[,index] ~ y_x[,index-1])$coefficients[[2]]
}
colnames(coeff_vecHMD) <- c("a","b")
# prepare the plots by getting data in df
library(tidyverse)
Year <- seq(1951, 2017) # generate a vector of ages from 0 to omega
Brass_df <- as.data.frame(cbind(Year=Year, alpha=coeff_vecHMD[,1], beta=coeff_vecHMD[,2])
```

```r
Brass_df = Brass_df %>%
  mutate(alphcma = zoo::rollmean(alpha, k = 6, fill = NA)) %>%
  mutate(alphatma = zoo::rollmean(alpha, k = 6, fill = NA)) %>%
  mutate(betacma = zoo::rollmean(beta, k = 6, fill = NA)) %>%
  mutate(betatma = zoo::rollmean(beta, k = 6, fill = NA, align = "right"))
# The plots
ggplot(Brass_df, aes(x=Year)) + theme_classic() +
  ggtitle("Trend in alpha parameter from 1951 to 2017") +
  geom_line(aes(y=alpha)) +
  geom_line(aes(y = alphcma, colour = "centred 6-yr ma"), na.rm=TRUE) +
  geom_line(aes(y = alphatma, colour = "trailing 6-yr ma"), na.rm=TRUE)

ggplot(Brass_df, aes(x=Year)) + theme_classic() +
  ggtitle("Trend in beta parameter from 1951 to 2017") +
  geom_line(aes(y=beta)) +
  geom_line(aes(y = betacma, colour = "centred 6-yr ma"), na.rm=TRUE) +
  geom_line(aes(y = betatma, colour = "trailing 6-yr ma"), na.rm=TRUE)
ILD <- read.csv("data/q_Ireland_by_Years.csv")
y_x <- 0.5*log((1-ILD[2:10])/ILD[2:10]) # logit transform q_x's
y_x <- rev(y_x) # reverse columns to put in chrono order: 1935-2015
# make linear model and gather a and b parameters
coeff_vecIL <- matrix(, nrow=9, ncol=2)
for (index in 2:length(y_x[1,])){
  coeff_vecIL[index-1, 1] <- lm(y_x[,index] ~ y_x[,index-1])$coefficients[[1]]
  coeff_vecIL[index-1, 2] <- lm(y_x[,index] ~ y_x[,index-1])$coefficients[[2]]
}
Year <- seq(1935, 2005, 10) # generate a vector of ages from 0 to omega
ILD_df <- as.data.frame(cbind(Year=Year, alpha=coeff_vecIL[1:8,1], beta=coeff_vecIL[1:8,2
ILD_df = ILD_df %>%
  mutate(alphcma = zoo::rollmean(alpha, k = 2, fill = NA)) %>%
  mutate(alphatma = zoo::rollmean(alpha, k = 2, fill = NA)) %>%
  mutate(betacma = zoo::rollmean(beta, k = 2, fill = NA)) %>%
  mutate(betatma = zoo::rollmean(beta, k = 2, fill = NA, align = "right"))
# The plots
ggplot(ILD_df, aes(x=Year)) + theme_classic() +
  ggtitle("Trend in alpha parameter from 1935 to 2005") +
  geom_line(aes(y=alpha)) +
  geom_line(aes(y = alphcma, colour = "centred 2-yr ma"), na.rm=TRUE) +
```

```
  geom_line(aes(y = alphatma, colour = "trailing 2-yr ma"), na.rm=TRUE)
ggplot(ILD_df, aes(x=Year)) + theme_classic() +
  ggtitle("Trend in beta parameter from 1935 to 2005") +
  geom_line(aes(y=beta)) +
  geom_line(aes(y = betacma, colour = "centred 2-yr ma"), na.rm=TRUE) +
  geom_line(aes(y = betatma, colour = "trailing 2-yr ma"), na.rm=TRUE)
```