Mathematical Analysis of Mortality Dynamics

by Dyfan Rhys Jones

13th, December, 2013

Contents

1 Introduction **3**
- 1.1 Gompertz Law of mortality data 4
- 1.2 Deviations from Gompertz Law 4
- 1.3 Models that incorporate Gompertz Law 5
- 1.4 Heterogeneous Population described by Gompertz Law 8
- 1.5 Mortality dynamics as a Poisson Process 10
- 1.6 Analysis of mortality data 10
  - 1.6.1 Likelihood and Least Squares 11
  - 1.6.2 Bayesian Information Criterion 14
- 1.7 Male and Female mortality data 14
- 1.8 Within the dissertation 15

2 Model fitting and Selection Criteria **16**
- 2.1 Model selection 16
- 2.2 Selection Criteria 18
- 2.3 Derivation of Bayesian Information Criterion (BIC) 19
- 2.4 Summary 25

3 Gompertz Law and Poisson Process **26**
- 3.1 Gompertz Law 27
  - 3.1.1 Evolution of Gompertz parameters 28
  - 3.1.2 Compensation Effect 29
- 3.2 Poisson Process 32
  - 3.2.1 Mortality data as approximated by function $x\, m_0 e^{\beta x}$. 34
  - 3.2.2 Mortality data as approximated by function $\frac{1}{x}\, m_0 e^{\beta x}$. 36
- 3.3 Summary 37

4 Heterogeneous Gompertz model **40**
- 4.1 Evolution of parameters 41
- 4.2 Reducing number of free parameters 43
- 4.3 Parameter predictions 46

4.4 Summary 46

5 Conclusions **49**

Chapter 1

Introduction

Over the last century people have come accustomed to high levels of risk. In the last century several major wars have swept across the globe; this would have of lead to people wanting to reduce the element of risk (in life in general). A method in reducing risk is insurance. Insurance companies have become the norm in today's society. It has even been made compulsory in certain parts of life, for example car insurance. This type of insurance attempts to reduce the risk of driving in a car. The main element of risk is a car crash, the insuring body will cover all costs incurred on the the vehicle.

There are many types of insurance that are optional. Insurance firms have taken advantage of every corner of daily life. As technology advances and becomes more portable insurance companies have taken it upon themselves to be the main method of reducing risk. A new form of insurance that has been developed recently is mobile phone insurance, this form of insurance is usually linked to devices people carry around with them on their person.

People have become so dependent on the idea of reducing the element of risk that insurance companies now insure the 'concept of life'. Life insurance is an attempts to compensate the recently deceased family/relatives with wealth in an attempt to remove the fear of lack finance the recently decease provided. For insurance companies to consider life insurance mortality dynamics will have to be taken into account. Without mortality dynamics the insuring body can run the risk of under estimating the total amount to insure a large group of clients.

When insurance companies tackle risk, they have to consider the two main types of risk; individual and aggregate. The first type of risk refers to random fluctuations around an expected value. This is down to the uncertainty of an individual's life. The other component is a systematic risk. This means that the risk is systematically deviating from the expected life time, due to unexpected mortality improvements [24]. The unexpected mortality has the most significant financial impact on insurance companies. With this in mind it is understandably obvious that research in mortality has become more of a priority in insuring bodies.

1.1 Gompertz Law of mortality data

The modeling of mortality over the past decades has attracted a growing body of researches [4][5]. A key development in the modeling of mortality occurred when Benjamin Gompertz published a paper in 1825 [3]. In his study he describes the link between mortality and aging. Benjamin Gompertz uses a exponential function for this link, it is more commonly known as *Gompertz's Law*. Mathematically Gompertz Law is expressed by;

$$m_x = m_0 e^{\beta x} \tag{1.1}$$

Where m_0 represents the initial mortality at age 0. β reflects the rate of demographic aging (or the change of mortality rate). x is the age of the population and m_x is the mortality relating to the specific age.

1.2 Deviations from Gompertz Law

The Gompertz function gives a good prediction (fitting) of mortality after a specific age [3]. This become apparent when the mortality against age is displayed on a graph.

By viewing mortality against age in a semi logarithmic scale, it becomes clear that mortality is split into three key stages. The first stage is a sharp decline in mortality between ages 0 to 10. This is down to the improvement of the immune system in youngsters at this specific aged period.

The second stage is called the accidental hump and occurs approximately around 12 to 30. The accidental hump can still affect the system even as late as the late 30's to early 40's. This stage reflects the accidents young adults have at this age. For females the accidental hump also takes into account menstrual deaths (deaths when giving birth).

The final stage is simply aging. This stage is where Benjamin Gompertz noticed the link between mortality and age. The increase in mortality is fairly linear in a logarithmic scale, and suggests an exponential link between morality and age after a specific age (approximately 40 plus). Pearson found the dividing of mortality into stages by considering deaths as a random event

[6]. He found out that mortality can potentialy be divided into five different stages. He used deaths from England (1871-1880) to show these stages.

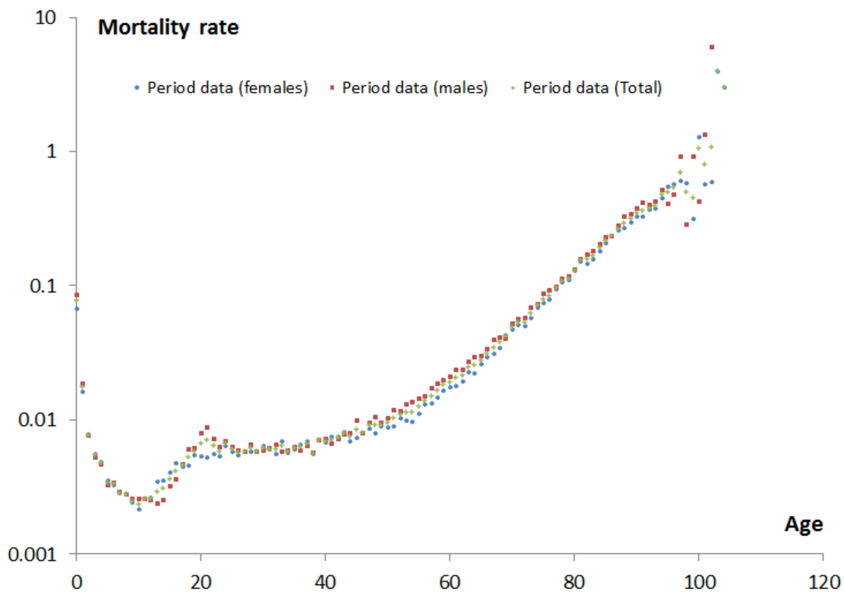

Figure 1.1: *Period data from Sweden 1915, set in a semi logarithmic scale (natural log of mortality against age). Consisting of female mortality (blue circles), male mortality (red squares) and total mortality (green diamonds).*

For the insurance companies the Gompertz function is ideal. It has demographic meaning for each parameter. It also has minimal number of free parameters. The Gompertz function is ideal for getting the trend of mortality over a time period. Through the development of trends predictions can be made for future time periods.

1.3 Models that incorporate Gompertz Law

Many models have incorporated the Gompertz function. The most noticeable model is the Makeham model [4].

$$m_x = \alpha + m_0 e^{\beta x} \tag{1.2}$$

The single difference is that this model includes a constant α. This constant has a greater effect depending on the species it is modeling mortality on. For example, for wild birds the α constant has a greater impact compared to a human population, the Gompertz function has a greater impact [27].

The addition of the parameter alpha has a certain act on the nature of the model. Firstly it increases the initial starting point at age 0. The second affect is that it makes the demographic slope be more gradual at early ages. This gives it the affect of a gradually increasing in gradient until it converts to the Gompertz function. When the Makeham model converts to the Gompertz function the parameter alpha has little affect to the overall model. From observing figure (1.2) it is clear to see the effect the alpha parameter has on the Makeham model. The initial increase of the parameter alpha from 0 to 0.0001 the model reveals a sharp jump from the initial mortality. As the alpha parameter continues to increase by 0.0001 the jumps between the previous model become small.

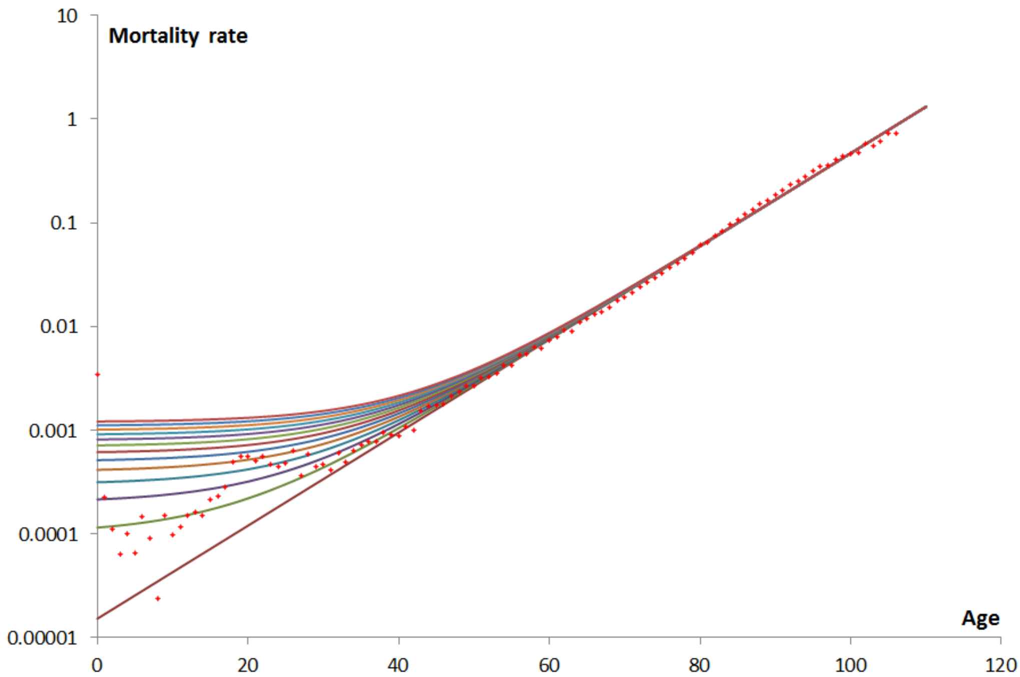

Figure 1.2: *Period data from Sweden 2000. Set in a semi logarithmic scale, the natural logarithm of mortality against age. The Makeham model has been fitted with constant initial mortality and beta. The alpha parameter is increasing from 0 to 0.0012 by a factor of 0.0001.*

Another model that incorporates the Gompertz Law is the Heligman Pollard Model [5] [24] [25]. This model attempts to fit the probability of dying over the probability of surviving. The model can be adapted to fit the force of mortality. Through observations and analysis in the pre-dissertations "Modelling Mortality Dynamics in Human Populations" mortality data and probability data have key similarities making the Heligman Pollard Model able to comparable against mortality data. Gompertz Law is within the third term of the Heligman Pollard model;

$$\frac{q_x}{p_x} = A^{(x+B)^C} + De^{-E(lnx - lnF)^2} + GH^x$$

Each term of the Heligman Pollard Model models each part of the mortality data. The first term models the data over the ages of 0 to 12 also referred to as the young age model. The Second term is takes into consideration the accidental hump and the final term fits the aging aspect of mortality dynamics.

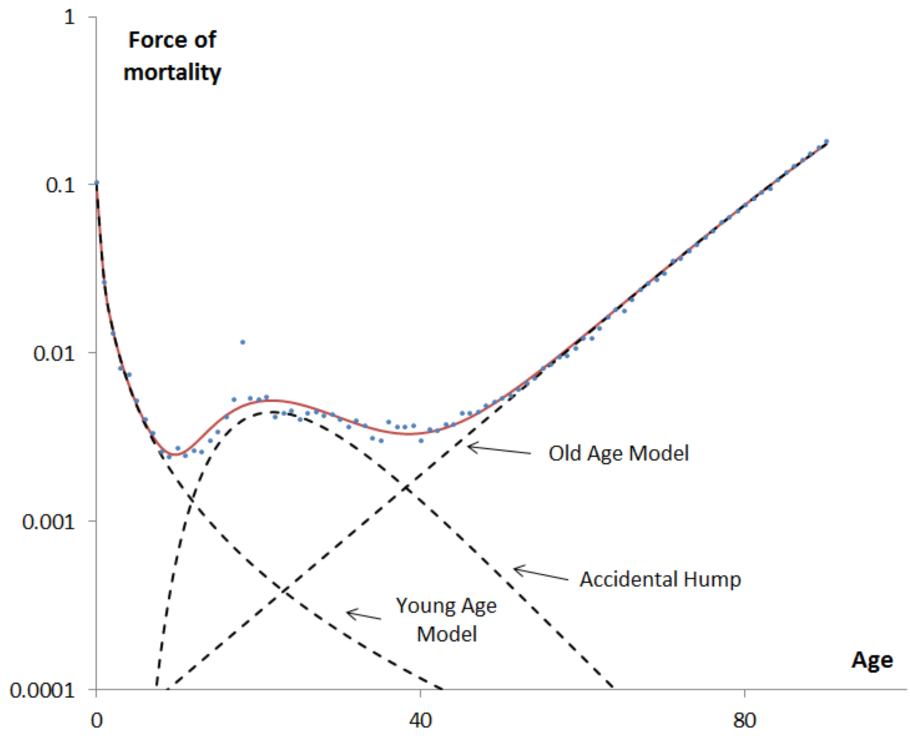

Figure 1.3: *Cohort data from Sweden 1900 capped at the age 90. Set in a semi logarithmic scale, natural logarithm of probability of dying against age.*

There are many other models that incorporate the Gompertz Law, but for this dissertation they won't be considered.

1.4 Heterogeneous Population described by Gompertz Law

In this stage we introduce another model of mortality which corporate the Gompertz Law and the Heterogeneity of human populations. Therefore for simplicity we refer to it in this dissertation by the name "Heterogeneous Gompertz model" [10]. This model has a simple idea, to break the population into sub-populations[8]. Each sub-population is then amenable to Gompertz function modeling over a specific point of mortality. This model suggests that the simple Gompertz function is the homogeneous version of the Heterogeneous Gompertz model. The Heterogeneous Gompertz model has several advantages over the simple Homogeneous Gompertz function; as each sub-population models a specific period of mortality (for example age 0 to age 5) the model can take form of the overall mortality.

Insurance companies will be more incline to want the information of a person's entire career life. If the assumption that people who pursue a career start in their early 20s (due to higher education), then the mortality at the age of 20+ can now be considered. By observing figure (1.1) it is clear that the accidental hump still affects the mortality of the population. This shows that the simple Homogeneous Gompertz function will be inadequate, because it will not take into account any affect the accidental hump has on the mortality. Also the Makeham model (equation 1.2) will be inefficient as this is only the Gompertz equation with an constant term added to it [4]. The Heterogeneous Gompertz model on the other hand will be ideal for it can take into account the tail end of the accidental hump. This is important because it will give a better level of accuracy. With the analysis tool of the Bayesian Information Criterion [16], the Heterogeneous Gompertz model can now determine the number of sub-populations [7][8][9]. As each sub-population will increase the number of k parameters, the Bayesian Information Criteria will give the higher sub-population models a high penalty. This will allow the number of sub-populations to stay low and to avoid the possibility of over fitting the data set.

The Heterogeneous Gompertz model [10] is represented in the following equation;

$$m_x = \frac{\sum_{j=1}^{n} \rho_{xj} \left(\frac{m_{xj}}{(1+(1-a)m_{xj})} \right)}{1-(1-a)\sum_{j=1}^{n} \rho_{xj} \left(\frac{m_{xj}}{(1+(1-a)m_{xj})} \right)} \tag{1.3}$$

where m_{xj} is the Gompertz function (equation 1.1) over the sub-population j. a is the fraction of the age where the average number of deaths occur within the year. As mortality is being considered a is typically set to half (0.5). Finally ρ_{xj} is the fraction of the total population for a specific sub-population;

$$\rho_{xj} = \frac{N_{xj}}{N_x}$$

where N_{xj} is the size of the sub-population and N_x is the size of the total population. By implementing the Gompertz function another parameter has to be taken into account (β_j). For each sub-population there are three corresponding parameters. By increasing the number of sub-population the number of parameters increases dramatically.

Due to the nature of the sub-population fractions, one of the free parameters can be removed. As the sum of the fractions of the sub-populations equal to 1 one fraction can be removed.

$$\rho_x = 1 - \left(\rho_{x1} + \rho_{x2} + \cdots + \rho_{x(j-1)}\right)$$

The Heterogeneous Gompertz model was firstly introduced in the paper "A mathematical Model of Mortality Dynamics across the life span combining heterogeneity and stochastic effects" [10]. Within the study the analysis of number of a sub-population in fitting mortality data was contemplated. To find the optimal number of sub-populations without over the fitting of mortality data Bayesian Information Criterion was used.

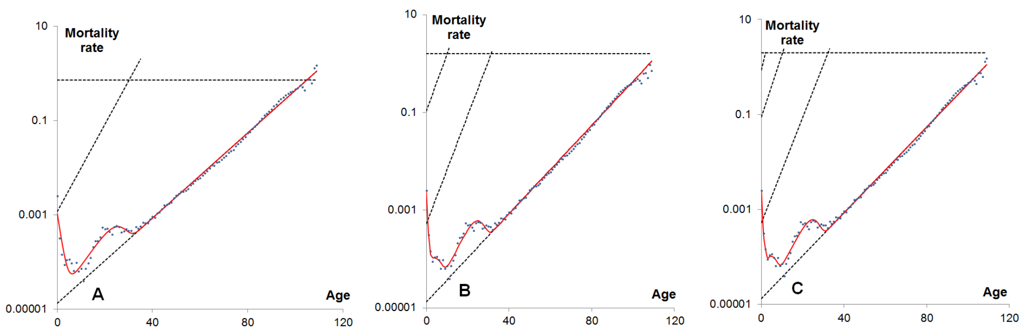

Figure 1.4: *Semi-Logarithmic scale, the logarithm of mortality against age. Fitting data from Sweden Mortality data in the year 2007. Graph A is a 3 sub-population Heterogeneous Gompertz model, Graph B is a 4 sub-population Heterogeneous Gompertz model and graph C is a 5 sub-population of Heterogeneous Gompertz model.*

Through the analysis using the Bayesian Information Criterion the 4 sub-population Heterogeneous Gompertz model is the best fitting model for mortality data. However it is difficult to deduce which sub-population corresponds to what stage of mortality data. The first sub-population model's the sharp decline in mortality data which is based on approximately age 0 to 10. The second sub-population model's the accidental hump that starts approximately age 12 and finishes roughly at age 25. The third sub-population reflects the aging aspect of mortality. Now the four sub-population improves the fit of the young mortality giving a better fit. The fifth sub-population points to the improvement of the fit at the end of the accidental hump. However the fifth sub-population starts to make the Heterogeneous Gompertz model to over fit the mortality data.

1.5 Mortality dynamics as a Poisson Process

Gompertz Law has been used as a key method in modeling mortality after the age of approximately 40, however Gompertz Law can also be derived by the Poisson Process [11].

$$P(n,x) = \frac{1}{n!}(\lambda x)^n e^{-\lambda x}$$

The Poisson Process takes the Poisson distribution and applies it to the idea of mortality. To do this the assumption that deaths can by caused be illness and disease. It is noted that probability of death by illness or diseases are randomly independent so illness or disease are not caused by a prior illness or disease and they do not influence an illness or disease.

There are two different types of the Poisson Process, the homogeneous and the non-homogeneous [12]. The homogeneous Poisson Process is one of the most well known processes. This process focuses on the rate of parameter λ, also known as intensity parameter. The non-homogeneous Poisson Process takes the idea that the rate parameter is changing in respect to time. So the generalized rate parameter can be expressed as a function $\lambda(f)$. For this dissertation the Homogeneous Poisson Process will be taken as a method to derive the Gompertz Law and also express additional possible functions that could fit the mortality data.

1.6 Analysis of mortality data

The analysis of data is an important stage of statistical mathematics. There are many methods for statistical analysis, many are base on assumptions over the overall data set. These assumptions consist of using the probability density functions as a method to form statistical analysis. Many of these methods of analysis are called hypothesis test [13] and contain the main common forms;

- z-test statistic: This test uses the normal distribution to form an analysis.
- t-test statistic: This test also uses the normal distribution to form statistical analysis however this test considers a small sample size compared to the z-test statistc.

With the use of these hypothesis testing a large range of analysis can be formed. However there is a large range of data sets that do not form hypothesis test and therefore they do not take the probability density function approach. This approach attempts to fit a model to the data set. Linear regression is a a typical model that is a commonly used in fitting data sets. The most basic form of linear regression is the following;

$$y_i = \beta_0 + \beta_1 x_i + \xi_i$$

where β_0 is the cross section of the y-axis, β_1 is the gradient of the slope, x_i is the explanatory variable $(i = 1, \ldots, n)$. ξ_i is the error of the linear regression [14].

However when trying to fit models to a data such as the linear regression there are two key methods, the Likelihood and the Least Squares [15].

1.6.1 Likelihood and Least Squares

A method of analysis need to be developed [17]. For this dissertation Bayesian Information Criterion (BIC) will be the main method for the analysis. BIC can be represented into two formats. The likelihood format and the least squares format. Each format gives the same result showing a casual key link between the likelihood and residual sum of squares. To understand this link the Likelihood and Least Squares theory will have to be considered.

Typically Least Squares theory is favoured over Likelihood Theory. Biologists have been exposed to the Least Squares theory through the application of applied statistics, as a result Least Squares theory has ‘enjoyed’ an early history of application (Weistedburg 1985) [17]. By comparison Fisher’s Likelihood Methods have required numerical methods and have turned out to be unpopular before the development of the computer. Both the Least Squares and the Likelihood have many similarities [15]. They both yield identical estimators of the structural parameters for linear and non linear models, this is when the residuals are assumed to be independent and normally distributed [23].

To see the link between Likelihood and Least Squares theory, the normal distribution $N(\mu, \sigma^2)$ can be considered. This is where μ is the mean of the normal distribution and σ^2 is the variance

$$g(x_j|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\frac{(x_j-\mu)^2}{\sigma^2}}.$$

The likelihood of the the probability distribution is expressed as;

$$L(\underline{\theta}, x_j) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\frac{\sum_{j=1}^{n}(x_j-\mu)^2}{\sigma^2}}$$

$$L(\underline{\theta}, x_j) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n e^{-\frac{1}{2}\frac{\sum_{j=1}^{n}(x_j-\mu)^2}{\sigma^2}}$$

The next step is to take the log-likelihood;

$$ln(L(\underline{\theta}, x_j)) = ln\left(\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n e^{-\frac{1}{2}\frac{\sum_{j=1}^{n}(x_j-\mu)^2}{\sigma^2}}\right)$$

$$= ln\left(\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n\right) + ln\left(e^{-\frac{1}{2}\frac{\sum_{j=1}^{n}(x_j-\mu)^2}{\sigma^2}}\right)$$

$$= n\ ln\left(\frac{1}{\sqrt{2\pi}}\right) - n\ ln(\sigma) - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - \mu)^2$$

The maximum likelihood estimator (MLE) needs to be found. To do that partial derivatives of the log-likelihood of the normal distribution. By setting it to 0 the maximum likelihood can be formed and further analysis can be observed. The partial derivation of the log-likelihood of the normal distribution in respect to μ is the following;

$$\begin{aligned} 0 &= \frac{\partial}{\partial\mu} l(\mu, \sigma^2; x_j) \\ 0 &= \frac{\partial}{\partial\mu} \sum_{j=1}^{n}(x_j - \mu)^2 \end{aligned} \tag{1.4}$$

$$
\begin{aligned}
0 &= -2\sum_{j=1}^{n}(x_j - \mu) \\
\hat{\mu} &= \frac{1}{n}\sum_{j=1}^{n}(x_j),
\end{aligned}
$$

and solving this for μ we can get an estimate for the population sample $\hat{\mu} = \bar{x}$. This can now be subituted into the maximum likelihood in respect to σ.

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \sigma} l(\mu, \sigma^2; x_j) \\
0 &= \frac{\partial}{\partial \sigma}\left(n\ ln\left(\frac{1}{\sqrt{2\pi}}\right) - n\ ln(\sigma) - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - \bar{x})^2\right) \\
0 &= -\frac{1}{\sigma}\left(\frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j - \bar{x})^2 - n\right) \\
\hat{\sigma}^2 &= \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2
\end{aligned}
$$

Through maximum likelihood the Least squares method has be shown. In addition to the maximum likelihood deriving the least squares method through the use of the normal distribution another important result can be expressed.

$$
log\left(L(\hat{\underline{\theta}})\right) = n\ ln\left(\frac{1}{\sqrt{2\pi}}\right) - n\ ln(\sigma) - \frac{1}{2\hat{\sigma}^2}\sum_{j=1}^{n}(x_j - \bar{x})^2
$$

or

$$
log\left(L(\hat{\underline{\theta}})\right) = -\frac{n}{2}ln(\hat{\sigma}^2) - \frac{n}{2}ln(2\pi) - \frac{n}{2}
$$

The additive constants can often be removed from the log-likelihood, this results to a key result.

$$
log\left(L(\hat{\underline{\theta}})\right) \approx -\frac{n}{2}ln(\hat{\sigma}^2)
$$

This result is important for the Bayesian Information Criteria, because it allows a simple mapping from Least Squares analysis results into the maximized value of the log-likelihood function.

1.6.2 Bayesian Information Criterion

Bayesian Information Criterion can be expressed in two forms giving it a greater power of analysis[17][19]. Firstly the likelihood format;

$$BIC = -2log\big(L(\hat{\underline{\theta}})\big) + k\,log(n) \tag{1.5}$$

The Least Squares format of the Bayesian Information Criteria assumes that all errors are set to a normal distribution and with a constant variance.

$$BIC = n\,log(\hat{\sigma}^2) + k\,log(n) \tag{1.6}$$

where k is the number of parameters and n is the number of data.

From this the models can now be analysed in greater detail. The Bayesian Information Criteria also takes into account parameters. This shows that the Bayesian Information Criterion gives a high penalty to the number of parameters (k). This suggests that models with fewer parameters will be favoured over models will larger number of parameters [20][21].

1.7 Male and Female mortality data

From observing figure 1.1, the number of sub-populations needed to model age 20+ is unclear. The Bayesian Information Criterion can be used to help to decipher the amount of sub-populations needed to model mortality above the age of 20+.

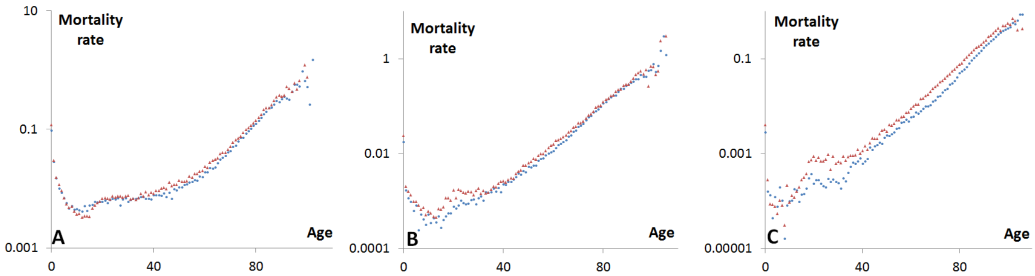

Figure 1.5: *Period data from Sweden, set in a semi logarithmic scale (natural logarithm of mortality against age). Male population is represented by the red triangles and the Female population is represented by the blue circles. Panel A is population from the year 1900, panel B is population from the year 1950 and finally panel C is population from the year 2000*

By observing male and female mortality through the last century there is an increasing gap between the two sexes. Figure 1.5 demonstrates the increasing difference between male and female mortality. The females' accidental hump seems to be decreasing in magnitude and with the period data set of 2000 it appears that this has disappeared. The decline in the accidental hump for female mortality could be down to the increase in medical advances in the field of giving birth. Due to these improvements male and female mortality in Sweden have grown steadily apart making it impractical to consider both sexes individually. As a result of this the combined mortality of Sweden will taken for the main analysis for this dissertation. For this dissertation all human mortality observation data is gathered from the website *"http://www.mortality.org"*.

1.8 Within the dissertation

For this dissertation there are a few key points that will be taken into consideration. The first point is the see if the Gompertz function is an adequate model for insurance companies to use when mortality of a population over the age of 40 is considered. This runs along side the second point the Poisson Process being another candidate for the insurance companies to use on the population. The comparison between Poisson Process and Gompertz Law can be analysed by the Bayesian Information Criterion. The next point focuses on the understanding of the BIC. This dissertation will go in depth into analysing how the BIC is derived and what assumptions it uses for the comparison of models.The final point is to consider a population over the age of 20 (as this will be more beneficial for the insurance companies due to nature of employment). For this the Heterogeneous Gompertz model can be used to make predictions for the future mortality. Also the reduction of free parameters will be explored, in an attempt to improve the predictions of the Heterogeneous Gompertz model's parameters.

Chapter 2

Model fitting and Selection Criteria

2.1 Model selection

In Statistics models are generally being developed to form an analyses over a data sample. The use of models allows a greater understanding of the data sample. However the choice of a suitable model is difficult, the possibility of over fitting or under fitting the data sample is high. By choosing an inappropriate model for the data sample the predictions and results can be drastically inaccurate. To avoid this from happening Model Selection has been developed so that a model can be analysed with regards to it's suitability for the data sample. To do this analysis several considerations have to been taken into account.

When finding a suitable model from a row of possible candidate models the number of parameters that the govern each model will have to be taken into account. The number of parameters of a model show whether the model has a chance in over fitting the data sample. Usually the larger the number of parameters will impact on the model over fitting the data sample. If the model has a small amount of parameters then it generally will under fit the data sample. This means that the model is too rigid to fit the data and will make poor predictions. Similarly if the model is over fitting a data sample then the model will only make predictions to the specific details of the data sample There is also the danger that the model will be fitting the noise within the data sample.

The other consideration that is taken into account when selecting a suitable model for the data sample, is the size of the sample itself. Generally speaking the increase in the sample size can compensate for the increase in the parameters of the model. However if a model can fit a large data sample efficiently with a small number of parameters this will be highly preferred. Due to the minimum number of parameters the model will have a fairly simple form, this will allow an easy analysis and predictions.

There are several methods for choosing the best model for the data sample from a group of candidate models. Most of these methods take into account the number of parameters within the model and the data sample size. This makes it difficult to choose which model selection method to use. There are a few categories for selecting a model. Each category holds a wide range of model selection methods. Here a few categories of model selection;

- Bayesian Model selection.
- Frequentest Model selection.
- Multi-level Inference: A Unifying view of Model selection [16].

The two noticeable model selection categories come straight from two aspects of statistics. Bayesian statistics is a widely used aspect of statistics. The key feature of this statistics is the use of priors linked with data to produce a posterior outcome. This feature runs throughout Bayesian statistics. The other noticeable model selection category (Frequentest Model selection) is linked to Frequentest statistics. This statistics does not taken into account any form of priors to form a result from. The frequentest approach is to use solely data provide with in the test. The final category takes a slightly different approach it considers multiple levels of inference where each level corresponds to one set of parameters or hyper-parameters. As each category of model selection is equally valid as the next, Bayesian Model selection will be considered for the analysis for this dissertation.

The reason for the use of Bayesian Model Selection will become more apparent when the assumptions are taken into account. With Bayesian model selection there are many different methods, each using slightly different assumptions for the selection of a model to fit the data. The two main model selecting approaches is the Akaike Information Criterion (AIC) [22] and the Bayesian Information Criterion (SIC, BIC and SBC) [17]. For this dissertation the Bayesian Information Criterion will be considered to be the main approach in comparing models given a data set.

2.2 Selection Criteria

The Bayesian Information Criterion was introduced by Schwarz G in 1978 in his paper 'Estimating the Dimension of a Model' [18]. Schwarz's criterion became a competitor to the already existing method of model comparison Akaike Information Criterion. Schwarz derived Bayesian Information Criterion as an approximation to the transformation of the Bayesian Posterior Probability for a candidate model. The derivation of Bayesian Information Criterion is mainly concerned with the log-likelihood and does not need particular requirements for priors allowing vague priors to be used within the derivation.

Another Bayesian comparison method is the Bayes factors [20]. The Bayes factors represents the ratio of the posterior probabilities of two candidate models with probable priors. The only problem with Bayes factors is that the priors' of the models can be difficult to set. Fortunately Bayesian Information Criterion model selection is approximately equivalent to the Bayes factors' model selection (in certain settings) [20]. This makes Bayesian Information Criterion a more favourable approach to model selection over Bayes factors when the priors are difficult to set.

Another criterion is the Akaike Information Criterion. Both criteria have very similar formulas with only part of the formal that differs.

$$\begin{aligned} BIC &: nln(\sigma^2) + kln(n) \\ AIC &: nln(\sigma^2) + 2k \end{aligned}$$

The difference between these two criteria is the second term [17]. The second term reflects on the penalty the criterion gives to the parameters of the candidate model. The penalty term for the Bayesian Information Criterion is more significant compared to the Akaike Information Criterion. As the sample size becomes greater than 8, the Bayesian Information Criterion penalty term as a greater impact to the criterion compared to the Akaike information Criterion penalty term. This makes the Bayesian Information Criterion favour models that have less parameters compared to the other criterion.

As each criterion is very close to each other it is difficult to choose between them. The Akaike Information Criterion is asymptotically efficient however it is not consistent where as the Bayesian Information Criterion is consistent yet not asymptotically efficient. A criterion that is asymptotically efficient will asymptotically select the fitted candidate model which minimises the mean squared error of prediction [17]. The Bayesian Information Criterion can be used to compare candidate models that are non-nested. The Bayesian Information Criterion can also be used to compare candidate models based on different probability distributions. By doing this the log-likelihood form of the Bayesian Information Criterion (equation 2.9) cannot discard any constant terms.

2.3 Derivation of Bayesian Information Criterion (BIC)

In deriving the Bayesian Information Criterion there are many methods. For this dissertation only one method for deriving the criterion will be considered. To start the derivation of the Bayesian Information Criterion lets consider a crude derivation of the criterion, so that the true derivation will become easier to understand.

Firstly lets consider an observed data sample $Y_n = (y_1, y_2, \ldots, y_n)$ and the data sample is n large [19]. Now lets assume that a candidate model M fits the observed data sample. The candidate model has k parameters that can be represented as a vector θ.

Now let $L(\theta|Y_n, M)$ represent the likelihood for Y_n based on the candidate model M. As shown in the introduction chapter the maximum likelihood estimator can be formulated from the likelihood. The maximum likelihood estimator for the candidate model is represented as $L(\hat{\theta}|Y_n, M)$. The next step is to take the marginal likelihood, this is represented as the following [18];

$$\int L(\theta|Y_n, M)d\theta \tag{2.1}$$

Now the likelihood can be approximated through Taylor's series expansion $\left[f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots\right]$. The second order of the Taylor's series expansion of the log-likelihood about $\hat{\theta}$ will be a efficient approximation of the log-likelihood function represented as $log\,(L(\theta|Y_n, M))$.

$$\begin{aligned} log\,(L(\theta|Y_n, M)) &\approx log\left(L(\hat{\theta}|Y_n, M)\right) + (\theta-\hat{\theta})^T \frac{\partial\, log\left(L(\hat{\theta}|Y_n, M)\right)}{\partial\theta} \\ &+(\theta-\hat{\theta})^T \frac{\partial^2 log\left(L(\hat{\theta}|Y_n, M)\right)}{2\partial\theta_i\partial\theta_j}(\theta-\hat{\theta}) \end{aligned}$$

The first derivative of the maximum likelihood estimator $\left(L(\hat{\theta}|Y_n, M)\right)$ is equal to zero. This makes the second term zero, the Taylor series expansion of the $logL(\theta|Y_n, M)$ can be reduced to;

$$\begin{aligned} log\left(L(\theta|Y_n, M)\right) &\approx log\left(L(\hat{\theta}|Y_n, M)\right) \\ &+(\theta-\hat{\theta})^T \frac{\partial^2 \, log\left(L(\hat{\theta}|Y_n, M)\right)}{2\partial\theta_i \partial\theta_j}(\theta-\hat{\theta}) \end{aligned} \tag{2.2}$$

The second derivative of the log-likelihood is a key component. The Fisher information corresponds to this component. The negative expected value of the second derivative of the log-likelihood is Fisher information. To understand this component a better understanding of the Fisher information will have to be considered and analysed.

Side note: Consider the normal distribution as an example in how the Fisher information works [23]. Suppose a random sample $(X_1, \ldots, X_n)$ from a normal distribution $N(\mu, \sigma^2)$, and μ and σ are unknown parameters. The Fisher information is represented by $I(\theta)$, where θ represents the unknown parameters μ and σ^2.

Firstly consider a single data point X_1. The Fisher information then can be just the log of the normal distribution.

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(X_1-\mu)^2}{2\sigma^2}}$$

$$l(x|\theta) = log\left(N(\mu, \sigma^2)\right) = -\frac{1}{2}log(2\pi) - \frac{1}{2}log(\sigma^2) - \frac{(X_1-\mu)^2}{2\sigma^2}.$$

Hence getting the second derivative $l(x|\theta)$ in respect to μ.

$$l_\mu(x|\theta) = \frac{(X_1-\mu)}{\sigma^2}, \qquad l_{\mu\mu}(x|\theta) = -\frac{1}{\sigma^2}$$

Now to get the negative expected value of the second derivative.

$$I_{\mu\mu} = -E\left[-\frac{1}{\sigma^2}\right] = \frac{1}{\sigma^2} \tag{2.3}$$

The second step is to get the second derivative $l(x|\theta)$ in respect to σ^2.

$$l_{\sigma^2}(x|\theta) = \frac{(X_1-\mu)^2}{2\sigma^4} - \frac{1}{2\sigma^2}, \quad l_{\sigma^2\sigma^2}(x|\theta) = -\frac{(X_1-\mu)^2}{\sigma^6} + \frac{1}{2\sigma^4}$$

Hence,

$$I_{\sigma^2\sigma^2} = -E\left[-\frac{(X_1-\mu)^2}{\sigma^6} + \frac{1}{2\sigma^4}\right] \tag{2.4}$$

The expected value for X_1 and X_1^2 are known and can be substituted into (equation 2.4) so that the negative expected value for the second derivative in represent to σ^2 can be found. Where $E[X_1] = \mu$ and $E[X_1^2] = \mu^2 + \sigma^2$ the Fisher information value is;

$$I_{\sigma^2\sigma^2} = \frac{1}{2\sigma^4}. \tag{2.5}$$

Finally the second partial derivative in respects to μ and σ^2;

$$l_{\mu\sigma^2}(x|\theta) = l_{\sigma^2\mu}(x|\theta) = -\frac{(X_1-\mu)}{\sigma^4}, \quad I_{\mu\sigma^2} = 0.$$

The Fisher Information values can now be placed into a matrix to form the Fisher's information matrix ($\bar{I}_1$ represents the Fisher's information matrix for one observation) for the normal distribution.

$$\bar{I}_1 = \begin{pmatrix} I_{\mu\mu} & I_{\mu\sigma^2} \\ I_{\mu\sigma^2} & I_{\sigma^2\sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}. \tag{2.6}$$

This matrix is now called Fisher's Information Matrix. A key point to take into account is the size of the Fisher's information matrix. The Fisher's information matrix is a k by k sized matrix where k is the number of free parameters. In this case there are two unknown parameters so the size of the Fisher's Information matrix is 2 by 2. Now to consider the log-likelihood of the normal distribution for the n data points.

$$l(x|\theta) = log\left(N(\mu,\sigma^2)\right) = -\frac{1}{2}log(2\pi) - \frac{1}{2}log(\sigma^2) - \frac{\sum_{p=1}^{n}(X_p-\mu)^2}{2\sigma^2}$$

where $p = 1,\ldots,n$. By taking the likelihood of the normal distribution all possible values of X are being considered. This will have a slight affect on the Fisher's information

$$
\begin{aligned}
l_{\mu}(x|\theta) &= \frac{\sum\limits_{p=1}^{n}(X_p)-n\mu}{\sigma^2}, \qquad l_{\mu\mu}(x|\theta) = -\frac{n}{\sigma^2} \\
l_{\sigma^2}(x|\theta) &= \frac{\sum\limits_{p=1}^{n}(X_p-\mu)^2}{\sigma^4} - \frac{1}{2\sigma^2}, \quad l_{\sigma^2\sigma^2}(x|\theta) = -\frac{\sum\limits_{p=1}^{n}(X_p-\mu)^2}{\sigma^6} + \frac{1}{2\sigma^4} \\
l_{\mu\sigma^2}(x|\theta) &= \quad l_{\sigma^2\mu}(x|\theta) = \quad -\frac{\sum\limits_{p=1}^{n}(X_p)-n\mu}{\sigma^4}
\end{aligned}
$$

By applying the negative expected value on each second derivative the Fisher's information matrix ($\bar{I}_n$ represents the Fisher's information matrix for the entire data set) becomes the following;

$$
\bar{I}_n = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix} = n\bar{I}_1. \tag{2.7}
$$

This characteristic of the Fisher's Information matrix carries through when finding the Fisher's information matrix for any model. The normal distribution was considered as the model in calculating the Fisher's information matrix as it is fairly easy to compute and analysis.The Fisher's information matrix is the inverse of the covariance of the normal distribution, this allows the covariance to be used when deriving the Bayesian Information Criterion.

Now lets substitute in the Fisher's information matrix into equation 2.2

$$
\begin{aligned}
log\,(L(\theta|Y_n, M)) &\approx log\left(L(\hat{\theta}|Y_n, M)\right) \\
&\quad -(\theta-\hat{\theta})^T \frac{1}{2} n\bar{I}_1(\theta-\hat{\theta}) \\
L(\theta|Y_n, M) &= L(\hat{\theta}|Y_n, M)e^{-\frac{1}{2}(\theta-\hat{\theta})^T n\bar{I}_1(\theta-\hat{\theta})} \\
where \qquad n\bar{I}_1 &= -E\left[\frac{\partial^2\, log\left(L(\hat{\theta}|Y_n, M)\right)}{\partial\theta_i\partial\theta_j}\right]
\end{aligned}
$$

The covariance can now replace the Fisher's information matrix and substituted into equation 2.1. This gives the following;

$$
L(\hat{\theta}|Y_n, M)\int e^{-\frac{1}{2}(\theta-\hat{\theta})^T n\bar{V}_1^{-1}(\theta-\hat{\theta})}d\theta \tag{2.8}
$$

The integral of the multivariate normal distribution is the following;

$$\int (2\pi)^{-k/2} |n\bar{V}_1^{-1}|^{1/2} e^{-\frac{1}{2}(\theta-\hat{\theta})^T n\bar{V}_1^{-1}(\theta-\hat{\theta})} d\theta = 1$$

Through the use of the integral of the multivariate normal distribution the expression can be simplified into an approximation. By taking the natural log and multiplying the approximation by -2 the Bayesian Information Criterion can be expressed.

$$\begin{aligned} & \int L(\theta|Y_n, M)d\theta \\ \approx \; & L(\hat{\theta}|Y_n, M)(2\pi)^{k/2}|n\bar{V}_1^{-1}|^{-1/2} \\ = \; & -2log\left(L(\hat{\theta}|Y_n, M)\right) - \frac{k}{2}log(2\pi) + klog(n) + log\left(|\bar{V}_1|\right) \end{aligned} \quad (2.9)$$

To get the Bayesian Information Criterion into the form that it is more commonly known then the terms that do not contain the number of observed data can be discarded. This discarding of constant terms happens when n becomes very large or tends to infinite, this makes the other terms within the Bayesian Information Criterion become insignificant.

$$BIC = -2log\left(L(\hat{\theta}|Y_n, M)\right) + klog(n)$$

This derivation of the Bayesian Information Criterion is only taking into account one possible candidate model for the data set. It also does not take into account priors, this is unusual as Bayesian statistic uses priors and data to get the posterior result. To consider the full derivation for the Bayesian Information criterion the priors and multiple candidate models will have to be taken into account.

If we assume that M_V be the candidate model that describes the data sample Y_n, the candidate model is selected from sequence of candidate models $M_1, \ldots, M_V, \ldots, M_L$ $(1 \leq V \leq L)$. The corresponding parameter vector for the candidate model M_V is represented as θ_V. in a similar fashion also to the previous proof of Bayesian Information Criterion $L(\theta_V|Y_n, M_V)$ is the likelihood for the candidate model M_V. The maximum likelihood estimator is also represented as $L(\hat{\theta}_V|Y_n, M_V)$ [19][21].

The priors can now be consider so that the Bayesian statistics can be formed. Assume that $\pi(M_V)$ denote a discrete prior [20] over the set of the candidate models $M_1, \ldots, M_L$. $\pi(M_V)$ is also assumed to have a positive probability. The prior over the parameter vector θ_V is denoted as $g(\theta_V|M_V)$. By applying Bayes' theorem to the above elements the posterior for M_V, θ_V can be worked out.

$$\begin{aligned} &\quad\; Posterior = \quad Prior \quad \times \quad likelihood\, of\, model \\ f(M_V, \theta_V|Y_n) &= \frac{1}{H(Y_N)} \quad [\pi(M_V)g(\theta_V|M_V)] \quad L(\theta_V|Y_n, M_V), \end{aligned} \tag{2.10}$$

where $H(Y_n)$ is the marginal distribution of Y_n. The Bayesian model selection procedure can then be based on choosing the model M_V which is the most probable posterior [21].

$$P(M_V|Y_n) = \int L(\theta_V|Y_n, M_V)d\theta_V. \tag{2.11}$$

The equation 2.11 can now be applied into equation 2.10.

$$P(M_V|Y_n) = H(Y_n)^{-1}\pi(M_V)\int L(\theta_V|Y_n, M_V)g(\theta_V|M_V)d\theta_V. \tag{2.12}$$

The applied Bayes' theorem for the most probable posterior is similar to the marginal likelihood equation 2.1. Now the priors of the candidate model and the corresponding parameter vector can now be non informative $\big(\pi(M_V),\ g(\theta_V|M_V) = 1\big)$. Finally the term $H(Y_n)^{-1}$ is a constant and for the purpose of model selection the term can be discarded. This now makes the probable posterior take the form of;

$$P(M_V|Y_n) \equiv S(M_V|Y_n) = \int L(\theta_V|Y_n, M_V)d\theta_V$$

This makes it exactly the same format as equation 2.1. However as we have considered different candidate models and priors the BIC is now non-bias and follows Bayesian statistics. The Bayesian Information Criterion be formulated in the exactly same method demonstrated before, in the crude definition.

$$BIC = -2log\left(L(\hat{\theta}_V|Y_n, M_V)\right) + klog(n).$$

This shows that the Bayesian Information Criterion does not need an informative prior. However informative priors can be used to derive the Bayesian Information Criterion, but non-informative priors make the derivation of the criterion more tractable.

2.4 Summary

From the 'Derivation of the Bayesian Information Criterion' section then BIC is shown to be an adequate method of selecting a model over a group of possible candidate models. It has also shown how the BIC does not need an informative prior to be derived. This allows the BIC to be a better method for model selection compare to Bayes' Factors[20].The Bayes' Factors need informative priors to make an informative decision on the best possible candidate model. As several model priors are difficult to calculate the BIC provides a better form of model selection.

As both AIC and BIC are in the Bayesian Model selection category it is difficult to choose between the two criterion. However the BIC has a strong penalty for the number of parameters, therefore BIC will be used for this dissertation.

Chapter 3

Gompertz Law and Poisson Process

For insurance companies the age group above the age of 40 is particular important, as this age group takes into account the retirement of the general population and finally deaths. For life insuring companies the age group that considers death is particular important for this will help to predict the deaths of the general population.

There are two suitable models that can fit this data group. Gompertz Law (also known as the Gompertz function) has be recognised as suitable method in model mortality data since the late 1800s. It has been known that the Gompertz Law will show the relationship between mortality and age above a certain age. However in reality mortality data in young and old data ages will deviate from the Gompertz function. A second suitable model is the Poisson Process, this approach is very new. The Poison process has the recognition that the probability of death can be caused by an premature illness.

Both models will be able to fit the mortality data above the age of 40. However for insurance companies the question is which model is best. To answer this question the Bayesian Information Criterion can be used to evaluate the two methods of fitting the mortality.

Another point of interest is the relationship between the parameters of Gompertz law. The relationship between the parameters can be observed to see if it is consistent within the Poisson Process. The Poisson Process is a key feature in this analysis as it can take may forms; because of these multiple forms the Poisson Process can take it is a virtual observation in whether the parameter relationship is constant.

3.1 Gompertz Law

Since the publication of the paper "On the Nature of the function Expressive of Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies" by Benjamin Gompertz [3], Gompertz Law has been one of the main forms of modeling mortality data. It has been so important that many models have incorporated the function within themselves.

The Gompertz Law contains two parameters that have demographical importance. The first parameter is the initial mortality (also referred as the theoretical mortality). The second parameter is the demographic slope of the Gompertz function (beta).

$$m_x = m_0 e^{\beta x} \tag{3.1}$$

where x represents age.

The link between the two parameters was first observed by Bernard L. Strehler and Albert S. Mildvan [29]. The link they observed is that the two parameters have an inverse relationship; high value of initial mortality is related to small value of demographic slope parameter β and vice-versa. However this link was formed with the use of the Makeham Model. By taking mortality data from above the age of 40 the Gompertz Law's parameters link can be observed.

By taking the Gompertz function of the last century and over laying the functions on top of each further analysis can be formed. This form of analysis has been used by L. A. Gavrilov and N. S. Gavrilov [27]. This analysis was called the Convergence of Gompertz functions and was observed across many different countries. The new observation will be taken from a single country over the last century. This will be more appropriate for insurance companies as the mortality rates of difference countries are different. Also insurance companies will be more inclined to consider data from the country they are situated in.

3.1.1 Evolution of Gompertz parameters

The first step for the insurance companies is to consider the mortality for older part of the population. After the age of 40, the general population will have been in employed for approximately 18 years. This suggests above the age of 40 this population has accumulated a significant amount of wealth, making it a priority for the insurance companies. By observing figure (1.1), the age group of 40 plus consists of the stage of aging, which has an exponential behaviour. This makes it perfect to fit the Gompertz function over this section.

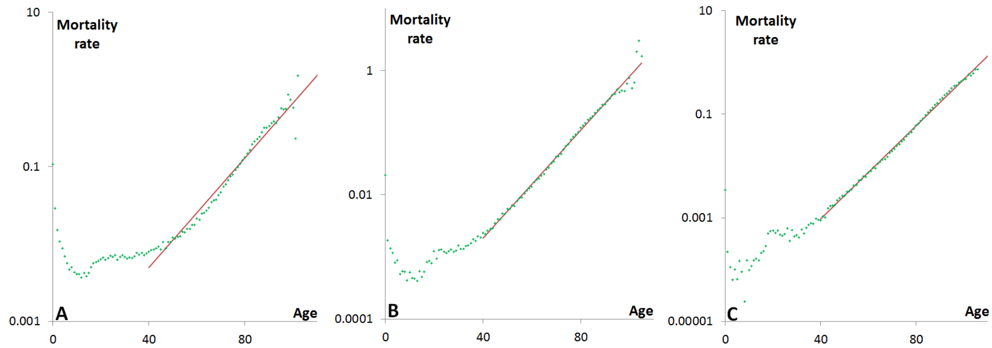

Figure 3.1: *Period data from Sweden, set in a semi logarithmic scale (natural logarithm of mortality against age). Each graph is fitted with the Gompertz function after the age of 40. Panel A is a population from the year 1900, panel B is a population from the year 1950 and panel C is a population from the year 2000*

By fitting the Gompertz function on the population of Sweden over the past century several observations and analysis can be formed. By observing figure (3.1) it is clear that the Gompertz function improves it's fit through time. In 1900 the Gompertz function doesn't fit the data well between the age of 40 to 50. This is down to the impact that the accidental hump has onto the data. The accidental hump at the beginning of the century has a low magnitude but it affects a large range of ages. Whereas in the later part of the century the accidental hump has a large magnitude but has a small range of the affected age groups.

From the Gompertz function (equation 3.1) there are two parameters (m_0 and β).and to examine their relationship we take their values by fitting Swedish period data over the last century with the Gompertz function using the Least squares method. The evolution of the two parameters over time is shown in figure 3.2. From this observation the relationship of the initial mortality and beta can be formed. As the initial mortality decreases the beta

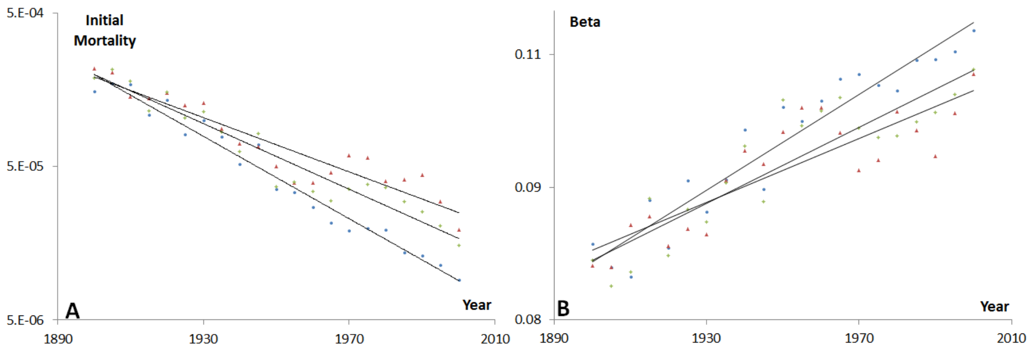

Figure 3.2: *Panel A plots the logarithm of mortality against years. The red triangles represent male population, the blue circles represents female population and finally green diamonds represent total population. Panel B Gompertz slope (Beta β) versus time set in years. Green triangles is the male beta, the blue circles represents the female beta and the green diamonds is the both male and female.*

parameter increases. This link is not exactly same rate. The initial mortality is decreasing at an exponential rate whereas the beta parameter is increasing at a linear rate. This link between the initial mortality and the parameter beta has been observed by Bernard L. Strehler and Albert S. Mildvan [29]. However in the study Bernard L. Strehler and Albert S. Mildvan have used the Makeham model (equation 1.2) instead of the Gompertz function on it's own. The link between the initial mortality and the Beta parameter is strong with both model designs.

3.1.2 Compensation Effect

Now to consider the link between the two parameters of the Gompertz function. Set the natural logarithm of the initial mortality at age 0 against the demographic slope beta (β), this gives a very surprising result. By observing figure (3.3) a clear link between these two parameters can be made. As the initial mortality at age 0 decreases exponentially the demographic slope increases linearly. This suggests as the initial mortality decreases the aging rate of the population increases.

A similar comparison has been made between the initial mortality and the demographic slope (β) with the studies [27] [29]. With the use of data that is taken from a large range of countries, the parameters are influenced by a large range of possible exterior effects. For example environmental factors may have impact on the data and also financial wealth of the country will also have an effect on the data. Whereas data from a single country over a

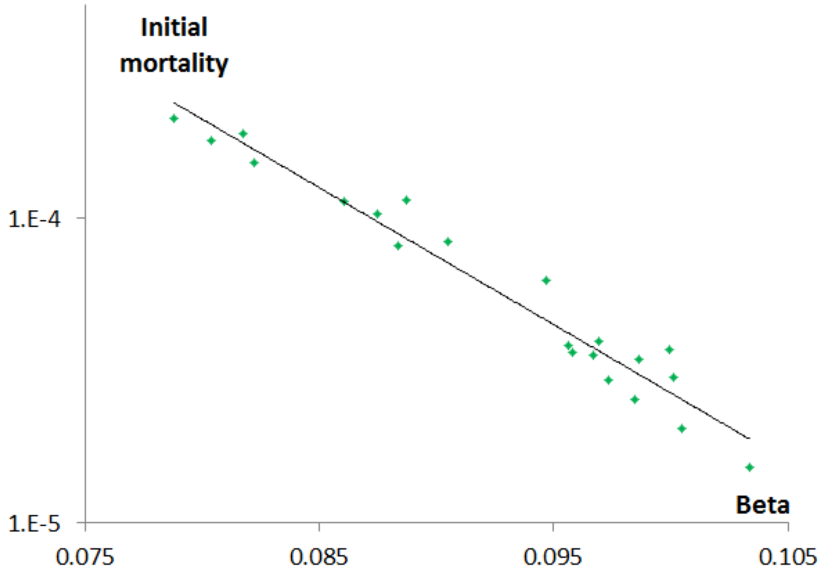

Figure 3.3: *Logarithm of the initial mortality rate versus Gompertz slope (β) at age 0 for the combined population of male and female for the country of Sweden.*

century, the parameters are only effected by the circumstances that happen within the country. As the environment of Sweden has not change over the last century, there will be no impact from the environment as it is constant. Sweden as a country has been classified as a first world country over the last century, illustrating that it is fairly wealthy and the general population from Sweden will have access to health care. From this it is clear that the two parameters will have little affect from environmental and financial factors. As there is little impact from exterior factors the compensation effect shown in figure 3.3 gives a near prefect analysis of the two parameters.

To get a better idea of the affect the two parameters have on the Gompertz function is to set several Gompertz functions that fit actual consecutive period data next to each other. Each Gompertz function will be representing a point in time over the last century. This approach of analysis is in the book published by L. A. Gavrilov and N. S. Gavrilova [27]. In the book the Gompertz functions that are set against each other, analyse data from different countries. This is slightly different from this dissertation, as previously mentioned the data gathered for this dissertation is from one country over one century.

Unlike the result shown in the book mentioned previously, two Gompertz functions stray out of the conclusion set by Gavrilov. The conclusion was

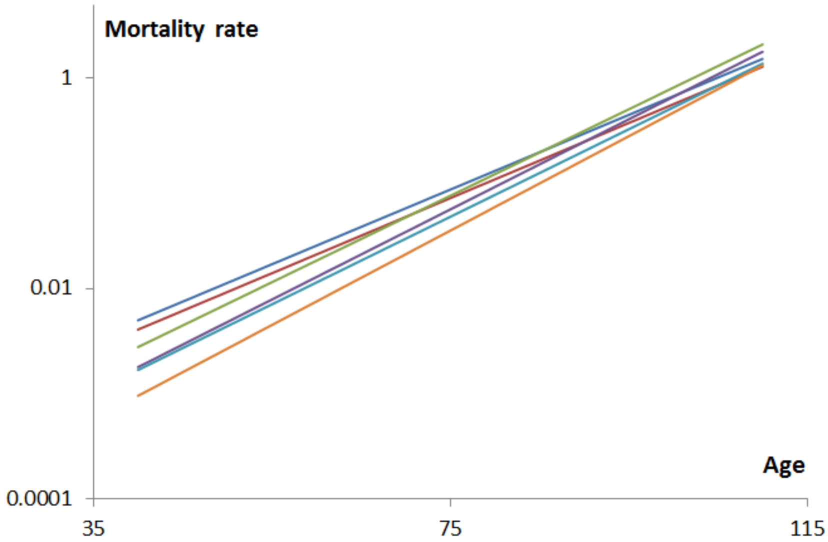

Figure 3.4: *Logarithm of mortality versus age over Swedish data. Dark blue that fits the 1900 Gompertz function, Maroon line is the 1920 Gompertz function, Green is the 1940 Gompertz function, Purple represents 1960 Gompertz function, Light blue is the 1980 Gompertz function and Orange is the 2000 Gompertz function*

that all the Gompertz functions will approximately converge to a single point. It is demonstrated by graphical displays within the book. However the data used for this conclusion seems have been bias. The data has been picked from a large range of countries. The years in which the data is extracted from seem to be varied, and suggesting that the data was hand picked to draw up the conclusion. On the other hand the data could be randomly chosen, as the year and the country have been randomly selected.

The result from figure 3.4 gives a slight convergence of the Gompertz functions. There are two Gompertz functions that diverge from this conclusion they are the 1940's and the 1960's Gompertz functions. This result cannot be ignored as a simple anomaly. It occurs over the entire time period of 20 year. Each Gompertz function seems to have been shifted to converge at a different point making it a significant result. This shows that there is not a convergence to a single point but just a general convergence of Gompertz functions. However we can conclude that the compensation effect is observed. As initial mortality decline the Gompertz slope is increasing.

3.2 Poisson Process

So far the Gompertz function has been proven to be an excellent method on fitting mortality data after the age of 40. However can the Gompertz function be derived and expressed through the idea of the Poisson Process? When deaths are considered to be the cause of death then the Poisson Process takes the following assumption;

- Death is caused by a form of illness or disease.
- All illnesses or diseases are random independent events

As the Poisson Process has been considered the probability density function is the starting point for this approach [11]. The probability that an individual have n diseases at age x is given by;

$$P(n,x) = \frac{1}{n!}(\lambda x)^n e^{-\lambda x}, \tag{3.2}$$

where λ is the rate of one disease in a unit of age, n is the number of diseases and x represents age. We consider that a fraction δ, average number of the diseases and therefore the force of mortality is given by;

$$\mu(x) = \sum_{n=0}^{\infty} \left[P(n,x)(\delta n) \right] \tag{3.3}$$

where δ the probability of deaths due to illness.

$$\mu(x) = \delta \sum_{n=0}^{\infty} \frac{1}{(n)!}(\lambda x)^n e^{-\lambda x}$$

let $e^{-\lambda x}$ be constant. The term $e^{-\lambda x}$ does not depend on n and can be taken out of the sum. The Taylor's expansion $\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$ can be used to simplify the expression further.

$$\begin{aligned} \mu(x) &= \delta e^{-\lambda x}(\lambda x)e^{\lambda x} \\ \mu(x) &= \delta(\lambda x) \end{aligned}$$

Therefore let's consider thatλ is a function of age x. The assumption that the rate λ is a constant give that the force of mortality is a linear function of age and this result is not very interesting. Then the force of mortality is given by;

$$\begin{aligned}\mu(x) &= \delta\sum_{n=0}^{\infty}\frac{1}{(n-1)!}(f(x)x)^{n-1}e^{-f(x)x}\\ &= \delta x f(x). \qquad (3.4)\end{aligned}$$

As proven by Benjamin Gompertz and observations, mortality behaves in an exponential format when mortality is against age. By taking a linear function for $f(x)$ it would prove to be uninformative, as mortality is exponential. Assume δ is constant for this study it will be set as 1. Now lets consider $f(x)$ as the Gompertz function, this will give;

$$\begin{aligned}f(x) &= \lambda_0 e^{\beta x}\\ \mu(x) &= x\lambda_0 e^{\beta x} \qquad (3.5)\end{aligned}$$

By changing the $f(x)$ function the Gompertz function can be found through this Poisson Process, when the function $f(x)$ is represented in the form;

$$\text{if } f(x) = \frac{1}{x}\lambda_0 e^{\beta x} \text{ then } \mu(x) = \lambda_0 e^{\beta x} \qquad (3.6)$$

Through the Poisson Process the Gompertz function can be derived. This makes the assumptions that the Poisson Process can possibly be assumed over the Gompertz function. This supports the Gompertz function being a suitable and good method of fitting mortality. By increasing the power of the denominator of equation 3.6 another format of the Poisson Process can be expressed.

As the Poisson Process can be used as a homogeneous model, it can be easily put against the Gompertz function for comparison. This will prove to be helpful for the insurance companies when selecting a candidate model. Unlike the Gompertz function the Poisson Process does not assume the function that fits the aging stage to be an exponential function.However through recent calculations the linear version of the Poisson Process gives an non-informative model when considering mortality.

Using the Poisson Process now can be represented into three different formats;

$$\begin{aligned}\text{if } f(x) = &\ \lambda_0 e^{\beta x} \text{ then } && \mu(x) = x\lambda_0 e^{\beta x}, \qquad (3.7)\\ \text{if } f(x) = &\ \tfrac{1}{x}\lambda_0 e^{\beta x} \text{ then } && \mu(x) = \lambda_0 e^{\beta x}, \qquad (3.8)\\ \text{if } f(x) = &\ \tfrac{1}{x^2}\lambda_0 e^{\beta x} \text{ then } && \mu(x) = \frac{1}{x}\lambda_0 e^{\beta x}. \qquad (3.9)\end{aligned}$$

where the parameter x represents age.

The first version of the Poisson Process will have a distinctive shape when set into a semi logarithmic scale graph. The function will give a convex shape that will fade gradually off as age increases. The second version of the Poisson Process (equation 3.8) is the Gompertz function and has been analysed in the previous section. On the other hand the third version of the Poisson Process (equation 3.9) will have a concaved shape that will smooth off into a straight line. These two shapes are very critical as either one could possibly fit the mortality against age, better than the Gompertz function.

3.2.1 Mortality data as approximated by function $x\, m_0 e^{\beta x}$

The first version of the function $f(x)$ when placed in the Poisson Process can be referred to "exponential multiplied by age" (where the exponential is the Gompertz function). As the Poisson Process can be modeled over a homogeneous population it makes sense to model the Poisson Process over the same age period as the Gompertz function. This will allow for further analysis between the two approaches in modeling mortality.

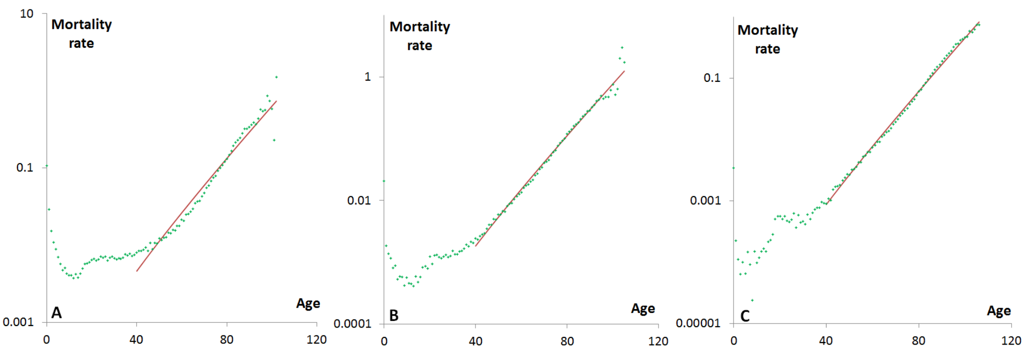

Figure 3.5: *Logarithm of mortality versus age (period data from Sweden). Each graph is fitted with the Poisson Process with the function given in equation 3.7 fitted after the age of 40. Panel A is the population at the year 1900, Panel B is the population at the year 1950 and finally Panel C is the population at the year 2000*

By observing figure (3.5) the fit of the Poisson Process improves over time similar to the improvements of the Gompertz function over time. The convex shape of the "exponential multiplied by age" Poisson Process does not quite match the general shape of the mortality above the age of 40. In the early part of the 20th century the general shape of the logarithmic mortality above the age of 40 is concaved for the first part and then assumes a straight line that is increasing due to age. By the end of the century the shape of the logarithmic mortality after the age of 40 is completely different, it becomes roughly a straight line that is increasing proportional to age. Taking this in to account, does the "exponential multiplied age" Poisson Process fit the mortality worse than the Gompertz function? This question can only be answered through therefore analysis using Bayesian Information Criterion. However before this analysis is made the comparison between the two parameter need to be observed.

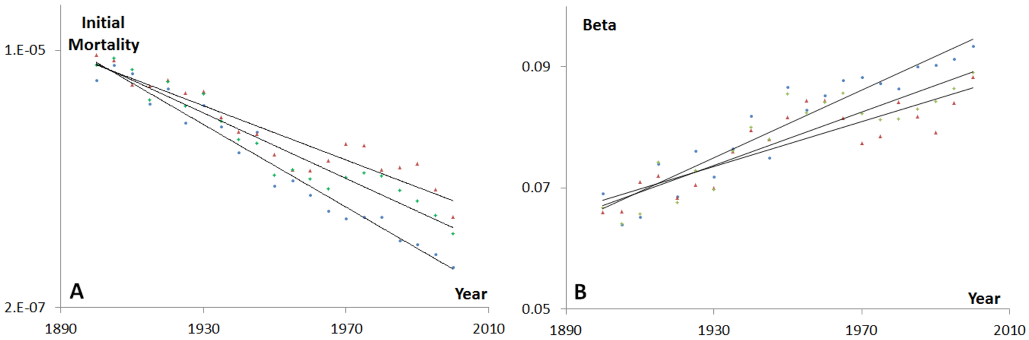

Figure 3.6: *Both panels are fitted with the function given in equation 3.9. Panel A has the logarithm of mortality against years. The red triangles represent male population, the blue circles represents female population and finally green diamonds represent total population. Panel B has the Gompertz slope (Beta β) versus time set in years. Green triangles is the male beta, the blue circles represents the female beta and the green diamonds is the both male and female.*

From figure (3.6) the parameters are acting exactly the same way as the Gompertz function. The parameters compared to the Gompertz function are ever so slightly smaller in size. However the link between the exponential decrease in the initial mortality and the linear increase of the (β) parameter still remains. These strengths reveal the unusual relationship between the two parameters first noticed by Bernard L. Strehler and Albert S. Mildvan with the use of the Makeham model (equation 1.2).

So far only "exponential multiplied by ages" has been observed and analysed, the next Poisson Process has not yet been consider. The next Poisson Process that will be introduced separately from the Gompertz function will be the "exponential divided by age" Poisson Process.

3.2.2 Mortality data as approximated by function $\frac{1}{x} m_0 e^{\beta x}$

In a similar fashion to the "exponential multiplied by age" Poisson Process the "exponential divided by age" Poisson Process will be fitted to data of age 40 plus. This will allow the comparison between all three methods in modeling for a homogeneous population.

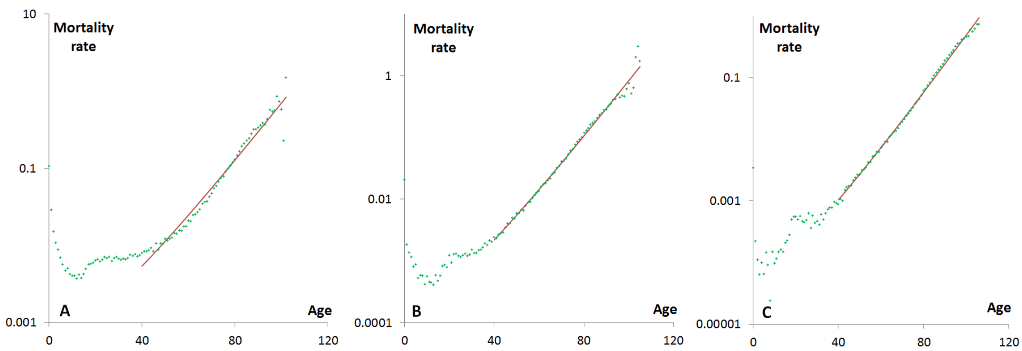

Figure 3.7: *Logarithm of mortality versus age (period data from Sweden). Each graph is fitted with the Poisson Process with the function given in equation 3.9 fitted after the age of 40. Panel A is the population from the year 1900, panel B is the population from the year 1950 and finally panel C is the population from the year 2000.*

Unlike the previous Poisson Process, the "exponential divided by age" has a strong similarity to the logarithmic mortality after the age of 40. This suggests that this version of the Poisson Process could prove to fit the mortality far better. The concaved structure of the "exponential divided by age" Poisson Process takes into consideration the accidental hump. From figure (3.7) a similar observation to the previous Poisson Process can be made. The "exponential divided by age" Poisson Process fits the mortality better later in the last century. As the Poisson Process now contains a different function ($f(x)$) the link between the two parameter could be very different from previous fitted models.

From observing figure (3.8) the parameters for the "exponential divided by age" Poisson Process is exactly the same as the previous Poisson Process and the Gompertz function. Through both methods of the Gompertz function and the Poisson Process the parameters (β and m_0) present the same pattern. As the initial mortality decrease exponentially the beta parameter increase linearly in respect to time. This reinforces the link between the two parameters.

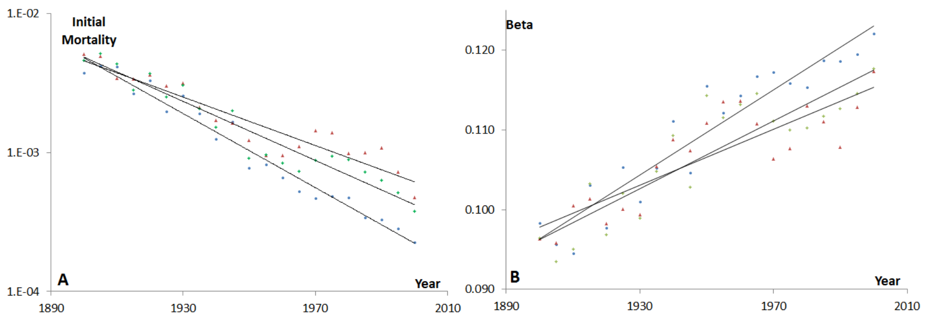

Figure 3.8: *Both panels are fitted with the function given in equation 3.9. Panel A is the logarithm of mortality against years. The red triangles represent male population, the blue circles represents female population and finally green diamonds represent total population. Panel B. is the Gompertz slope (Beta β) versus time set in years. Green triangles is the male beta, the blue circles represents the female beta and the green diamonds is the both male and female.*

3.3 Summary

Firstly the comparisons between the Poisson Process can be formed. Through the previous analysis of the parameters a strong link has been clearly been shown. As the initial mortality decreases the beta parameter increases linearly. This link is also present in the Gompertz function, showing that this link between the two parameters in constant. Similar to the Gompertz function analysis the two parameters can be set against each other (initial mortality versus beta in a semi-logarithmic graph) for further analysis for the Poisson Processes.

Like that of the Gompertz function the Poisson Process parameter act in the same manner. As the demographic slope (beta) increases the initial mortality decrease. The difference between these two Poisson Processes is the size of the parameters and not the trend between parameters. This corresponds to the shape of the Poisson Process. As the "exponential multiplied by age" Poisson Process has a convex shape, then initial mortality should be very small. With the demographic slope of this Poisson Process as apart the demographic slope is taken into account by the initial convex shape of the fitting model. As the "exponential divided by age" Poisson Process has a concaved shape then the initial morality should be slightly larger. Also the demographic slope (beta) will also be larger and this is down the to initial concaved shape of this Poisson Process. Finally as anticipated the Gompertz function is between the two Poisson Processes. Each compensation effect run parallel to each other in the semi logarithmic scale.

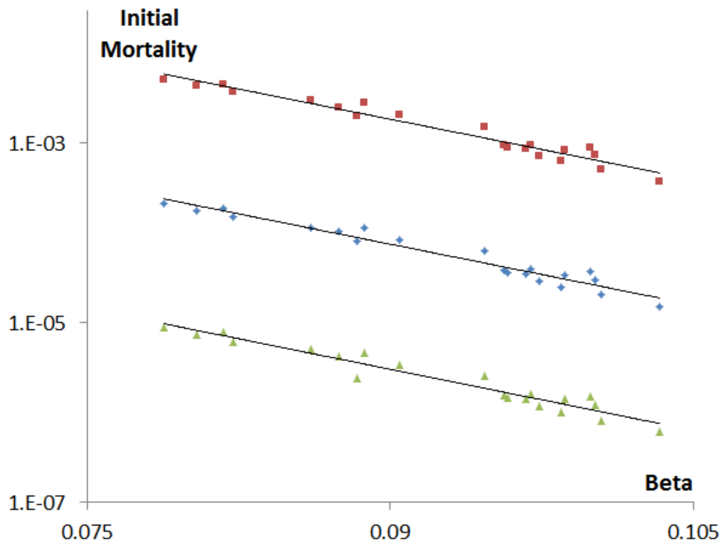

Figure 3.9: *Initial mortality versus Beta set in a semi-logarithmic graph, the black line is a fitted exponential function. The green triangles points taken by using the exponential multiplied by age, the blue diamond points is the Gompertz function. Finally the red square points is the exponential divide by age.*

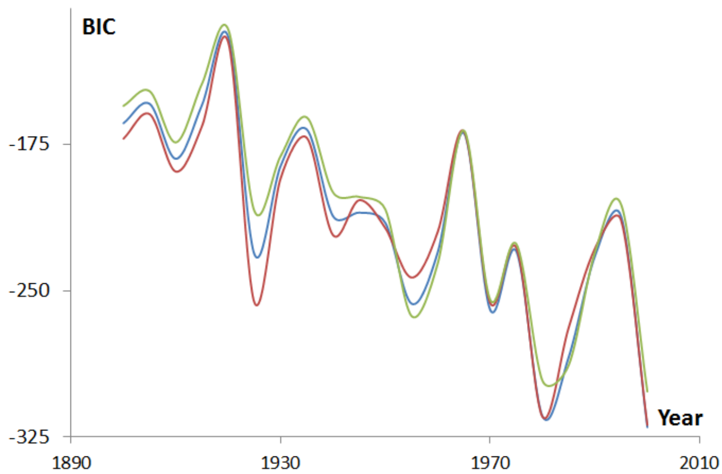

Figure 3.10: *The Bayesian Information Criterion (BIC) of all fitting models of mortality above the age of 40 against time. Where blue line represents the BIC for the Gompertz function, the red line represents the BIC for the "exponential divided by age " Poisson Process and the green line represents the BIC for the "exponential multiplied by age" Poisson Process.*

Secondly there is the comparison of modeling Poisson Process (both versions) against the Gompertz function. As the "exponential multiplied age" Poisson Process has a convex shape, this model would naturally be assumed to have the weakest fitting of mortality. From observing figure (3.10) the differences between the fitting models is very small. This makes it very difficult to choose the best fitting model. The Bayesian Information Criterion for all models follow the same pattern with little difference between each BIC. From figure (3.10) all models seem to be equally a good method for fitting mortality above the age of 40. When choosing which model to use for the fitting of mortality it all depends on the initial assumption on whether the probability of death is caused by an illness. When considering a diseases that can possibly kill individuals then the Poisson Process can be considered to be a better model. However the Poisson Process can be used to express the Gompertz function. The main method in picking the best fitting model will all depend on the population. Therefore it is unclear which method to use for analysis.

Chapter 4

Heterogeneous Gompertz model

After reviewing data above the age of 40 is it clear that the Gompertz function and the Poisson Process are efficient methods in modeling the data set. However insurance companies would prefer to take into account the point at which a person would start a career. This will allow for the wage structure of an individual's career to be assessed by the insurance company, making it far more beneficial for the insuring body in dealing with life insurance.

An assumption can be made in setting the point where the data is fitted by a model. As the entire career of an individual is now being taken into account it is reasonable to take data from the age of 20 and above. This assumption is that most people who start a career go through higher education and consequently postpone in taking a career until their early 20s.

By modeling the data from an earlier age the data now contains another section of the different stages of mortality [2]. To recap, the first stage is the sharp decline in mortality from the age of 0 to 10. This stage is still out of the range of the data set that is getting fitted by a model. The second stage is the accidental hump and this stage starts approximately from the age of 12 and finishes roughly at the age of 30. This stage is within the age group that is going to be used and fitted and analysed. However only half of the end part of the accidental hump stage is fitted and modeled. Fitting the Gompertz function and the Poisson Process maybe an inefficient method in fitting the data, and other modeling methods will have to be considered. The last stage of mortality is called aging, aging process is an exponential link between age and mortality.

By considering the Heterogeneous Gompertz model equation 1.3 as a suitable model to fit the age group of 20 plus, the model will be able to accommodate the remaining apart of the accidental hump. From the paper

‘A mathematical model of mortality dynamics across the lifespan combining Heterogeneity and stochastic effects’ (authors Avraam D, de Magalhaes J Pedro and Vasiev B) [10] the Heterogeneous Gompertz model was first derived and this demonstrated the versatility of the model. This versatility allows the model to take into account different aspects of the data set. This makes it an ideal model to use fit the age group of 20 plus.

By reflecting upon two sub-populations the Heterogeneous Gompertz model can use a Gompertz function to describe two stages of mortality that are happening over this age group. The first Gompertz function will describe the remaining part of the accidental hump and the second Gompertz function will portray the aging of mortality stage. These two sub-populations should be efficient enough to represent the data over the age of 20. By increasing the number of sub-populations can make the Heterogeneous Gompertz model start to over fit the data set. This is down to the increasing number of free parameters with each sub-population. However as a 2 sub-population will relate to the different stages of mortality within the data set, it would be more preferable to use a 2 sub-population.

When using the Heterogeneous Gompertz model for predictions the number of free parameters can cause problems. To reduce the number of free parameters without reducing the number of sub-populations, the free parameters can be fixed so that it cuts down the number of free parameters without diminishing the sub-populations.

4.1 Evolution of parameters

The evolution of parameters can be formed by fitting the Heterogeneous Gompertz model on several populations over the last century. By fitting the Heterogeneous Gompertz model on many populations over the last century several observations can be formed [24].

From figure 4.1 the Heterogeneous Gompertz model can be observed on how well it can fit mortality data over the age range. The 1900 period mortality data contains an accidental hump the spans over a large age range. The Heterogeneous Gompertz model takes this into account, the first sub-population described by the Gompertz function portrays the accidental hump up to the age of 40. The second sub-population then starts to take control of the model and moves the model from the accidental hump to the aging aspect of mortality. For the remaining time period are all represented in the same way. After 1950 the demographic slope of the first sub-population is decreasing, this could be a coincidence or it could down to affects of the accidental hump. The accidental hump starts to change shape from having a large span over an age group with a swallow curvature of the hump, to a short expanse of the accidental hump with a steep curvature of the hump.

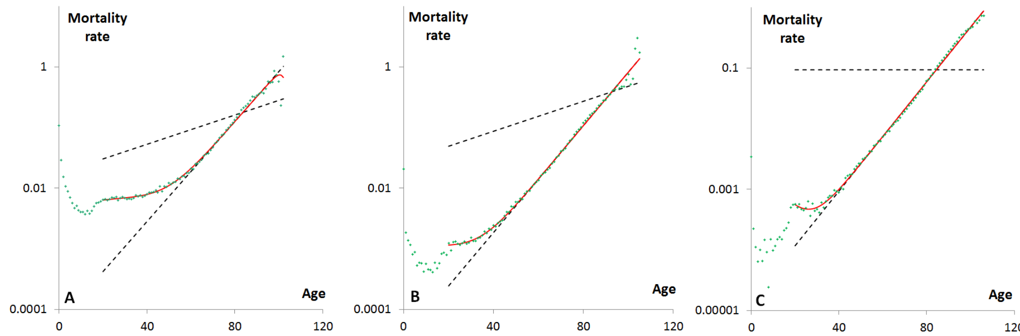

Figure 4.1: *Period data from Sweden, set in a semi logarithmic scale (natural logarithm of mortality against age). Each graph is fitted with the Heterogeneous Gompertz function after the age of 20. Panel A is the population from the year 1900, panel B is the population from the year 1950 and finally Panel C is the population from the year 2000.*

To see if the previous observations from figure 4.1 are accurate the parameters of the Heterogeneous Gompertz model can analysed. By viewing the parameters over last century trends and parameters can be observed to get a deeper understanding how the Heterogeneous Gompertz model changes throughout the last century.

By observing figure 4.2 the compensation effect appears. This compensation effect appears in the first sub-population. Through the advancements in medicine, the general concept for the initial mortality would be a decline in both sub-populations. This is not the case as observed in figure 4.2. The initial mortality for the first sub-population is increasing exponentially, this is very unusual. However it can be explained by the nature of the accidental hump. The accidental hump's steepness is fairly shallow at the start of the century, as time goes by the nature of the hump changes and by the end of the century the hump become steep. To compensate the change in the nature of the hump the initial mortality alters corresponding to the change. The second sub-population's initial mortality reflects the increasing improvements of medicine with in Sweden over the last century.

However the first sub-population can not have this effect as the initial mortality is increasing, the beta parameter does not display any form of trend or pattern instead it displays a random effect. Interestingly the parameters that reflect the fractions of the sub-populations is that one sub-population becomes more dominate as it increases size. This is suggesting that the

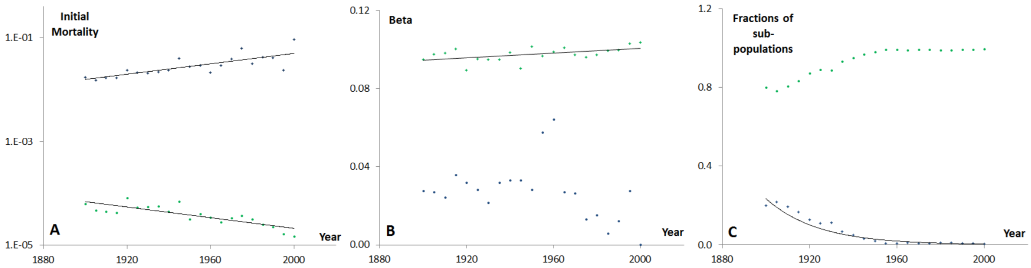

Figure 4.2: *For all graphs the blue points represents the first sub-population and the green points reflects the second sub-population. Panel A is the natural logarithmic of the initial Mortality versus the year of the period data. The green and blue points is fitted with an exponential trend. Panel B is the demographic slope against the year. Beta is fitted with an linear trend function. Panel C is the fraction of the sub-populations against year. The first sub-population is fitted with an exponential trend.*

data can increasingly be described by one sub-population, making it become approximately the Gompertz function.

4.2 Reducing number of free parameters

After observing the evolution of the parameters certain trends appeared that could be fitted with exponential and linear functions. When making any form of predictions it would be ideal if the number of free parameters can be minimised. When taking the Heterogeneous Gompertz model with 2 sub-populations there are 5 free parameters (the 6th parameter is not a free parameter). This gives countless of methods in reducing the number of free parameters. Now if the sub-populations within the Heterogeneous Gompertz model is taken into account, then the free parameters linking to the each sub-population can be use.

Between the two sub-populations it is the second sub-population that has largest fraction of the total population making it the most dominating sub-population. This is reflected in the fraction parameters from the previous section “Evolution of Parameters”. As a result the first sub-population’s parameters will be considered first. Each step in reducing the number of free parameters will make the model more rigid and consequently making the Bayesian Information Criterion number for the model poorer.

First Step: From the figure 4.2 the parameter from the first sub-population that is forming a trend is the initial mortality. By fixing the initial mortality parameter to exponential trend then the parameter is no long free and change the shape of the model ever so slightly making the parameters react accordingly.

Second Step: Through the fixing of the initial mortality of the first sub-population, the evolution of parameters change. The evolution of the beta parameter for the first sub-population there has been little improvement in setting a trend or pattern. Therefore the evolution of the parameter beta will be set to the average (a constant).

Third Step: Since the first sub-population's parameters have been fixed the second sub-population's free parameters can be taken into account. Both the initial mortality and the beta parameter of the second sub-population are showing trends. That makes this step fairly straight forward, the initial mortality of the second sub-population will be fixed to the exponential trend. This is leaving the evolution beta of the second sub-population being the final parameter to consider.

Fourth Step: The remaining beta parameter is now fixed to the linear trend. This makes all the parameters in relation to Gompertz function of each sub-population fixed. The only free parameter is one of the fractions of the population for the sub-population.The fractions are still presenting an exponential trend similar to the initial step from figure 4.2.

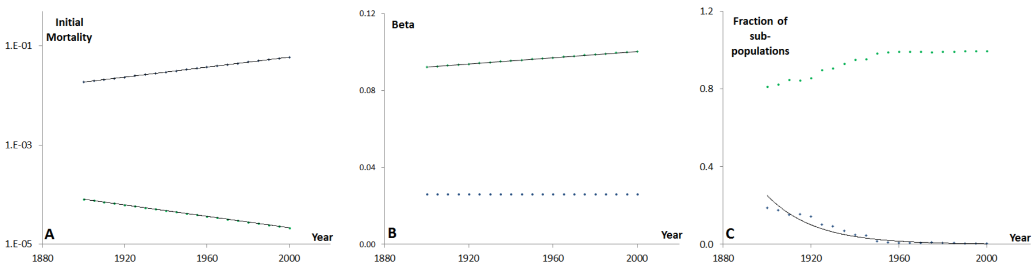

Figure 4.3: *For all graphs the blue points represents the first sub-population and the green points reflects the second sub-population. Panel A is in a semi-logarithmic graph of the initial Mortality versus the year of the period data. The green and blue points is fitted with an exponential trend. Panel B is the demographic slope against the year. Beta is fitted with a linear trend function. Panel C is the fraction of the sub-populations against year. The first sub-population is fitted with an exponential trend.*

Figure 4.3 is the evolution of parameters for the fourth step. By comparing the evolution of parameters of figure 4.3 with the initial evolution of parameters figure 4.2 there are strong similarities and there are some subtle differences. These subtle differences will have the biggest impact on how the Heterogeneous Gompertz model will change in shape over each step. Taking period data from the year 2000 the change in parameters with the fixing of parameters can now be shown in figure 4.4. The biggest change of the Heterogeneous Gompertz model is the first sub-population's Gompertz function. This is because of the beta parameter corresponding to this sub-population are being set to a constant.

By observing the fourth step to the initial step (step 0), there seems to be little change in the shape of the Heterogeneous Gompertz model. All the steps, taken into consideration, the accidental hump and morality are linked to aging. As there is not a great deal of change between the initial step and the fourth step, this suggest that the reduction in parameters may not be needed in this case.

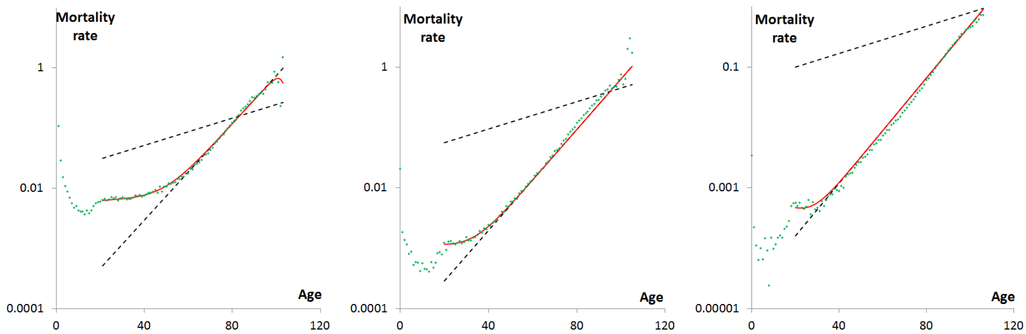

Figure 4.4: *Period data from Sweden from the year 2000, set in a semi logarithmic scale (natural log of mortality against age). Each graph is fitted with the Heterogeneous Gompertz function after the age of 20. Panel A is the population from the year 1900, panel B is the population from the year 1950 and finally Panel C is the population from the year 2000.*

4.3 Parameter predictions

To see whether the fourth step of the Heterogeneous Gompertz model will produce a better prediction of the parameters, it can be set against the parameters of the initial step of the Heterogeneous Gompertz model.

However unlike the prediction that the reduction of the free parameter could form a better prediction of future parameters, the reduced free parameters give a very similar result to the normal Heterogenous Gompertz model. This is disappointing but it can informative. The initial state of the Heterogeneous Gompertz model gives strong trends throughout most of the parameters (figure 4.2). Through the use of these trends the predictions for future parameters can be gathered making it suitable to consider the initial step without the unnecessary work to reduce the free parameters.

On the other hand the reduction of parameters is not always a necessary amount of work to obtain predictions. It can prove to be a very useful method for predictions.

4.4 Summary

From the observations of the evolution of parameters from the initial Heterogeneous Gompertz model (figure 4.2), the parameter that is linked to the fractions of the population presents some interesting results. The fractional parameters over time show that one of the sub-population becomes increasingly more dominant. The increasing in dominants starts to imply that is the Heterogeneous Gompertz model starts to become more like the Gompertz function. Suggesting that over time the data above the age of 20 can be represented by the removal of one of the sub-populations and therefore only the Gompertz function can be used.

From the previous chapter of "Poisson Process and Gompertz law" it has been shown that the Gompertz function can be derived through the Poisson Process. As a result of this the Poisson Process can be acknowledge for the comparison of the Heterogeneous Gompertz model. By taking this comparison the idea of the removal of one of the sub-populations can be considered.

The evolution of the fraction parameters from the initial step of the Heterogeneous Gompertz model (figure 4.2) at the start of the last century suggests that two sub-populations is going to be better than a single population. Which is shown by the first sub-population having approximately 20% of the total population at the start of the last century and by the end of the century the first sub-population has approximately 0.5% of the total population.

From figure 4.5 the first graph is displaying the Bayesian Information Criterion of all the models. The Poisson Process for the- 'exponential divided by age' is improving its fit most dramatically by the end of the century the BIC for Poisson Process 'exponent divide by age' become very close to the BIC of the Heterogeneous Gompertz model. This suggests that the Poisson Process 'exponential divided by age' for future period data sets could present an adequate model for insurance companies to consider.

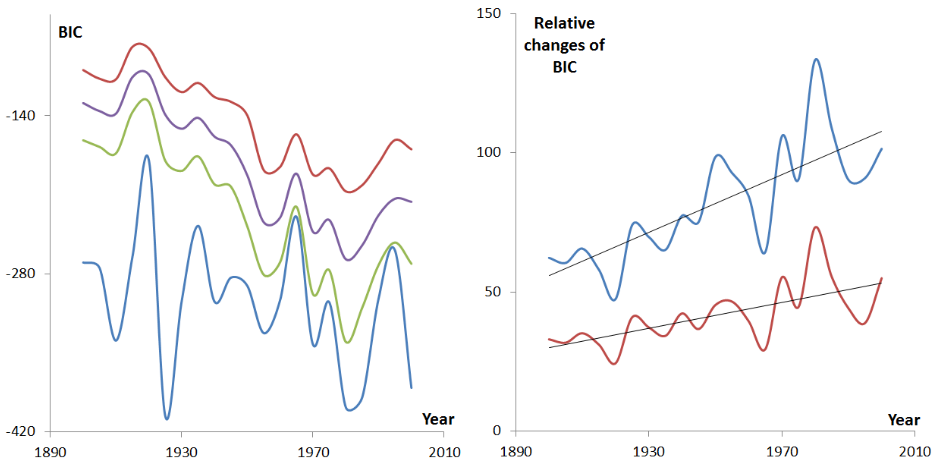

Figure 4.5: *Panel A is the red line is the BIC of the Poisson Process 'exponential multiplied by age', the purple is the BIC of the Gompertz function. The green line is the BIC of the Poisson Process 'exponential divided by age' and finally the blue line is the Heterogeneous Gompertz model with 2 sub-populations. Panel B shows the relative changes of BIC for different models. The blue line is the difference between the 'exponential multiplied by age' and the 'exponential divided by age'. The red line is the 'exponential divided by age' and the Gompertz function.*

Unlike the the comparison between all three of the Poisson Process above the age of 40, there is a clear difference the BIC values of each Poisson Process. By viewing the second graph in figure 4.5 the Poisson Process differences are displayed. The two differences made in the Poisson Process is the 'exponential divided by age' against the other two versions (Gompertz function and the 'exponent times by age').

The comparable difference between the 'exponential multiplied by age' and the 'exponential divided by age' is an increasing linear trend throughout the century. This is displaying that the 'exponential divided by age' is improving it's fit faster than the other Poisson Process.

The difference between the 'exponential multiplied by age' and the Gompertz function is also increasing by a linear trend. However it is increasing at a slower rate than the other difference.

When an insurance company is considering a model to use when fitting data above the age of 20, both the Heterogeneous Gompertz model and the Poisson Process ‘exponential divided by age’ version have to be contemplated. At the moment the Heterogeneous Gompertz model is still the best fitting model, however the Poisson Process ‘exponential divided by age’ is improving its fit at a steady rate. Giving it a possibility that the ‘exponential divided by age’ Poisson Process becoming as good as the Heterogeneous Gompertz model fit. But for now the main model to contemplate for fitting data above the age of 20 is the Heterogeneous Gompertz function.

Chapter 5

Conclusions

This dissertation is an extension of my prelimenary dissertation. Within the prelimenary dissertation the study of the Heterogeneous Gompertz model was reviewed and analysed. This dissertation extends the analysis of the Heterogeneous Gompertz model and introduces new model the Poisson Process. Also the BIC has been analysed in a greater extent enabling an greater understanding. Through the progress of this dissertation many factors about mortality dynamics have been addressed. The observation and analysis of this dissertation are listed as the following:

1. **Bayesian Information Criterion:** The comparison method of Bayesian Information Criterion which is the mathematical tool that is used to compare candidate models of a data set by taking into account the number of free parameters. The Bayesian Information Criterion can be defined in several ways. Each method with the definition of the criterion all start off with the initial step, taking the marginal Likelihood of the candidate model. To get to this step many methods consider placing the candidate model in the Bayesian theorem and then taking the probability of the posterior. Once set in this form different methods can be used to derive the criterion. The Laplace approximation of integrals is a common method in the deriving the criterion. Another method is to take the Taylor's expansion of the likelihood of the candidate model. Both methods in deriving the criterion aim to approximate the integral itself or apart of equation within the integral. As shown in the chapter "Model fitting and Selection Criteria" the derivation of the criterion does not depend on prior knowledge as it assumes a non-informative prior. It can also be derived in a crude method by the use of one candidate model instead of a set. The use of a set of candidate models has been the more common practice when deriving the criterion.

2. **Mortality data above the age of 40:** The first mortality data age group that was observed and analysed was the age group above the age of 40. When fitting the mortality data over this age group the Gompertz Law and the Homogeneous Poisson Process describing mortality dynamics were used as suitable models. Through the analysis of the Gompertz Law the compensation effect was applied to the Homogeneous Poisson Process. As the Gompertz Law was derived through the Homogeneous Poisson Process describing mortality dynamics (chapter "Gompertz Law and Poisson Process"). The parameters within the Gompertz Law are also present in the Poisson Process. The analysis of the different versions of the Poisson Process (equation 3.7, 3.8 and 3.9) clearly expressed the compensation effect within the parameters. This strengthened this effect in respect to mortality dynamics. When it came to comparing the Gompertz Law and the Homogeneous Poisson Process the Bayesian Information Criterion was used. From figure 3.10 the Bayesian Information Criterion number value of both the Gompertz Law and the Poisson Process was set against time. From this analysis neither modeling methods presented a clear 'winner'. As there was not a clear difference between the models.

3. **Mortality data above the age of 20:** The second mortality data age group that was observed was above the age of 20. With this the age group the Gompertz Law will be inefficient to model the mortality data. The reason for this was that the accidental hump is within the age group. The Heterogeneous Gompertz model was been used as a suitable model to fit the mortality data over the age group of 20 plus. The Heterogeneous Gompertz model presented the compensation effect over the second sub-population. This is the sub-population that is fitting the aging of mortality, which is approximately 40 and above. The first sub-population expressed an unusual result as the initial mortality did not decrease over time and instead it increased. However this could be explained by the nature of the accidental hump, the accidental hump became increasingly severe in height as time progressed. The initial mortality reflects the change in the nature of the accidental hump. The reduction of free parameters unfortunately did not produce a better prediction for the parameter compared to the initial Heterogeneous Gompertz model. By comparing the Bayesian Information Criterion value of the Heterogeneous Gompertz model and the Poisson Process over the same age group (above the age of 20), particular analysis and observation were made. The Poisson Process that is represented by the equation 3.9 shows the most promise in improvement over time. At the

end of the last century the Poisson Process's (equation 3.9) BIC value become very close to the Heterogeneous Gompertz model's BIC value. This presents a trend that the Poisson Process (equation 3.9) has the possibility in future populations can become equally as good as fit as the Heterogeneous Gompertz model above the age of 20.

4. **Poisson Process:** Through this dissertation the Poisson Process has been proven to be a suitable model for fitting mortality data. It can express Gompertz Law and alternative exponential functions. It is these functions that show promise as they are more flexible compared than the Gompertz Law, making it a key model for consideration when insurance companies analyses mortality dynamics.

Bibliography

[1] Samuel H. Preston, Patrick Heuveline and Michel Guillot, *Demography: Measuring and modeling population Processes*. Blackwell, Oxford 2000.

[2] Thomas R. Malthus, *An Essay on the Principle of Population*. Library of Economics and Liberty 1798.

[3] Benjamin Gompertz, *On the Nature of the function Expressive of Law of Human Mortality, and on a New Mode of Determing the Value of Life Contingencies*. Philos. Trans. R. Soc. Lond. 115 1825, p 513-585.

[4] Makeham,W.M., *On the law of mortality and the construction of annuity tables*. Assurance Magazine 1860, p 301-310.

[5] L.Heligman and J. H. Pollard, *The age pattern of mortality*. JIA 107 1980, p 49-80.

[6] Pearson, K., *The Chances of Death and Other Studies in Evolution*. Edward Arnold, London 1897.

[7] Vaupel, J.W., *Biodemography of human ageing*. Nature 426, 2010, p 536-542.

[8] Vaupel, J.W., Yashin, A.I., *The deviant dynamics of death in heterogeneous populations*. Sociol. Methodol., 1985, p 179-211.

[9] Vaupel, J.W., Manton, K.G., Stallard, E., *Impact of heterogeneity in individual frailty on the dynamics of mortality*. Demography 16 1979, p 439-454.

[10] Demetris Avraam, Joao Pedro de Magalhaes, Bakhtier Vasiev, *A mathematical model of mortality dynamics across the lifespan combining heterogeneity and stochastic effects*. Experimental Gerontololgy 48 2013, p 801-811.

[11] Ross S. M and Isham V. I, *Stochastic Processes*. Wiley. ISBN 9780-471-12062-9 1995.

[12] Snyder D. L and Miller *Random Point Processes in Time and Space*. Springer-Verlag. ISBN 0-387-97577-2 1991.

[13] David R. Anderson, Kenneth P. Burnham and William L. Thompson *Null Hypothesis Testing: Problems, prevalence and an alternative*. Journal of wildlife management 64(4) 2000, p 912-923

[14] P. J. Joseph, Kapil Vaswani and Matthew J. Thazhuthaveetil *Construction and Use of Linear Regression Models for Processor Performance Analysis*. The Twelfth International Symposium on High-Performance Computer Architecture. 2006.

[15] S. D. Pletcher, *Model fitting and hypothesis testing for age-specific mortality data*. J. Evol. Biol. 12 1999, p 430-439.

[16] Guyon I, Saffari A, Dror G and Cawley G, *Model Selection: Beyond the Bayesian/ Frequentist Divide*. Journal of Machine Learning Research 11 2010, p 61-87.

[17] Burnham, Kenneth P, *Model Selection and Multi-Model Inference : A Practical Information-Theoretic Approach (2nd Edition)*. Springer 2002.

[18] Schwarz G, *Estimating the dimension of a model*. Ann. Stats. 6 1978, p 461-464.

[19] Joseph E. Cavannaugh and Andrew A. Neath, *Generalizing the derivation of the Schwarz Information Criterion*. Communications in Statistics - Theory and Methods 28 1999, p 49-66.

[20] Kass R. and Raftery A, *Bayes Factors*. Jounral of the American Statistical Association 90 1995, p 773-795.

[21] Andrew A Neath and Joseph E. Cavannaugh, *The Bayesian Information Criterion: Background, deviation, and applications*. WIREs Computational Statistics 4 2012, p 199-203.

[22] Hamparsum Bozdogan, *Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions*. Springer, Psychometrika, Volume 52, Issue 3 1987.

[23] Songfeng Zheng, *Fisher Information and Cramer-Rao Bound*. http://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture

[24] Severine Gaille, *Forecasting mortality: when academia meets practice*. Eur. Actuar. J. 2012. p 49-76.

[25] Severine Gaille and Michael Sherris, *Modelling Mortality with Common Stochastic Long-Run Trends*. The Geneva Papers 2011. p 1-27.

[26] Thiel PN, *On a mathematical formula to express the rate of mortality throughout the whole life*. J Inst Actuar 1872, p 313-329.

[27] L. A Gavirlov and N. S. Gavrilova, *Determination of Species Length of life*. Dokl Akad Nauk SSSR 1979, p 465-469.

[28] L. A Gavirlov and N. S. Gavrilova, *Biology of Life Span: A Quantitative Approach*. CRC Press 1991.

[29] Bernard L. Strehler and Albert S. Mildvan, *General Theory of Mortality and Aging*. Science 132 1960, p 14-21.

[30] Anatoli I. Yashin, Ivan A. Iachine and Alexander S. Begun *Mortality modeling: A review*. Mathematical Population Studies 8 2000, p 305-332.