Australian National University

Advanced Age Mortality Modelling: an Extreme Value Theory Approach

Xu Ning

Supervised by Dr Fei Huang

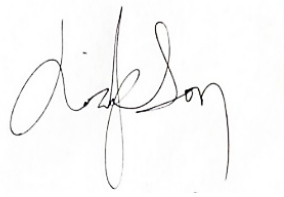
and Professor Ross Maller

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Actuarial Studies with Honours in 2019 at the Australian National University.

October 2019

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge and belief, contains no material published or written by another person, except where due reference is made in the thesis.



Xu Ning

October 31, 2019

Acknowledgements

My appreciation must first be given to my supervisor, Dr Fei Huang. Despite her insistence in treating me as a researcher of equal standing (much to my chagrin), there exists no positive probability I could have completed this project in the absence of her expertise and tremendous aid, which went far beyond the realm of what might be considered commonplace. In the same vein and with equal emphasis, I would like to thank Professor Ross Maller for co-supervising this project. His statistical expertise played a crucial part in reinforcing the arguments presented in this thesis; just as inspiring however was the way he changed my entire worldview up until now on what a professional document should look like. Perhaps he will forgive me for being flamboyant this one time. On a more serious note, I only have the utmost respect and gratitude for my supervisors and must emphasize my thanks to them for such a rewarding Honours' year, at the expense of their own precious time.

Special thanks to Nelson for help with many many things throughout the years, not least the wisdom he lent me for this Honours' thesis. Till the vicissitudes of fate do us apart.

Thanks also to Penghao for the plethora of information he has bestowed upon me in preparation for this year, and who knows how many years into the future. One paragraph dedicated to you, as promised.

'Most significant non-direct contribution to my thesis' prize goes to Kai for providing me with the legendary RPP's treasure. It will forever be immortalized in the glorious confines of my computer hard drive.

Finally, I must express gratitude for my family and friends for their unrelenting support in all matters physical and meta-physical. From those who laughed and cried in unison with me throughout this year, to those who were instrumental in making me what I am today: thanks for existing in this same snippet of space-time as me, against all the odds in the universe.

Abstract

In recent decades, the number of centenarians in developed countries has risen substantially as a result of improvements in quality of life and medical science (Robine and Caselli, 2005). The existence of this trend emphasizes the importance of accurately modelling extreme age survival probabilities. One recent development in this field is the threshold life table (TLT), due to Li, Hardy & Tan (2008). In the TLT model, from a sample or population of ages at death data, extreme value theory is used to make optimal use of the relatively small number of observations at high ages, while the traditional Gompertz distribution is assumed for earlier ages. In this thesis we use a high quality data set constructed by augmenting Netherlands population data from the Human Mortality Database (HMD) with a dataset provided by Centraal Bureau voor de Statistiek (CBS), or Statistics Netherlands. The Human Mortality Database is a joint initiative of the Department of Demographics at the University of California, Berkeley and the Max Planck Institute for Demographic Research. It provides detailed mortality and population data for 40 countries, including the Netherlands (Human Mortality Database, 2019). The HMD data set contains age and gender-specific death and number-alive counts from 1850-2016. The counts are given annually and thus interval censored; they are also right-censored at age 100. The CBS dataset consists of the ages at death (accurate to 1 day) of all Dutch residents aged 92 years or more whose year of death was between 1986 and 2015. To the combined data set as well as to the HMD data set alone, we fit two models: the threshold life table model as implemented in Li et al (2008), and a modified threshold life table model in which we incorporate a smoothing adjustment described in Section 3.4.2. From these models we estimate the highest attained age attained by members of the population or of cohorts within it and compare with the actual highest attained ages. We also compare the variance of fitted gammas and highest attained ages under the two models. We then use the smoothed threshold life table to value a theoretical annuity, and compare with valuations from the Heligman-Pollard as well as Coale-Kisker methods. Finally, we implement a dynamic adjustment for the smooth threshold life table and compare both the in-sample and out-of-sample sum of squared errors with those obtained from the Cairns-Blake-Dowd model.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background and Contribution

The number of supercentenarians in developed countries has increased considerably in recent decades (Vaupel and Robine, 2002). For life annuities and defined benefit pension plans, the rise of supercentenarians can lead to low-frequency, high severity losses (Li, Hardy, and Tan, 2008). Coupled with an increased awareness of longevity risk (Cairns, Blake, and Dowd, 2008) as well as the difficulty of reinsuring life annuities (Purcal, 2006), there has been a subsequent increased demand for actuaries to more accurately model extreme age survival probabilities.

However, the relatively less exposure to risk and consequent small numbers of deaths at high ages results in large sampling errors and highly volatile crude death rates, which makes modeling death rates at these ages difficult (Wilmoth, 1995). Furthermore, ages at death are often misreported, and substantial effort is needed to ascertain them accurately (Bernard and Vaupel, 1999; Bourbeau and Desjardins, 2002). In light of this, actuaries sometimes close life tables at an arbitrarily chosen age (eg. Coale and Kisker (1990)). Early closure however results in less accurate estimates of expected shortfall and value-at-risk for life annuity and defined benefit pension plan products, because these measures are by definition determined by the tail of the lifetime distribution (Duffie and Pan, 1997).

The threshold life table (TLT) method of Li, Hardy & Tan (2008) is a recent development that addresses these problems. In this model, from a sample or population of ages at death data, extreme value theory is used to make optimal use of the relatively small number of observations at high ages, while a traditional Gompertz law (Gompertz, 1825) is assumed for earlier ages. The threshold life table has two main attractive features. Firstly, it allows the end point of a life table to be determined in an objective data-driven way. Secondly, the threshold - the age at which the modelled lifetime distribution transitions from Gompertz to the distribution suggested by extreme value theory - is also selected objectively. However, it has a couple of drawbacks too. The model

is such that the death rate is not continuous at the threshold age, so the mortality transition from non-extreme to extreme ages is not smooth. Furthermore, the model structure is static and does not allow for forecasting of mortality rates into the future.

In this dissertation, we propose a smooth threhsold life table (STLT) model as an extension of the TLT model, which overcomes the continuity problem by incorporating a smoothing adjustment at the threshold age. Using a high quality dataset constructed by augmenting Netherlands population data from the Human Mortality Database (HMD) with a dataset provided by the Centraal Bureau voor de Statistiek (CBS) of the Netherlands, we demonstrate that the STLT not only guarantees a smooth transition of the death rate from non-extreme to extreme ages, but also provides a better fit to the data than the TLT model. We also use the STLT model to value a theoretical annuity, and compare with valuations from existing methods. Additionally, we introduce a dynamic adjustment for the STLT model and develop what we call the dynamic smooth threshold life table (DSTLT) model for forecasting. Empirical results using the augmented dataset show that the DSTLT model has superior in-sample fitting and out-of-sample forecasting performance compared to the Cairns-Blake-Dowd (CBD) model.

The rest of the dissertation is organised as follows. Section 2 provides a brief literature review of advanced age mortality models. Section 3 introduces the data sources used in this thesis and describes the construction of the augmented dataset. It then formulates the newly proposed STLT model and outlines the estimation methods. Section 4 presents the empirical results and analyses using both the proposed and existing methods. Section 5 outlines the construction of the DSTLT model based on results from fitting the STLT to a series of cohorts in the augmented dataset, and gives a comparison of it with the CBD model, cohort-by-cohort, both in and out-of-sample. Section 6 provides a discussion of findings. Section 7 concludes the thesis.

2 Literature Review

2.1 Notation

The following notation will be used throughout the thesis:

- X: age at death of an individual from the population, a continuous random variable
- $f(x)$: probability density function of X
- $F(x)$: cumulative distribution function of X
- $S(x) = 1 - F(x)$: survival function of F
- $h(x) = f(x)/(1 - F(x)) = -\frac{S'(x)}{S(x)} = -\frac{d}{dx} \ln S(x)$: force of mortality, or hazard function corresponding to F
- d_x: observed number of deaths in age interval $[x, x + 1)$
- E_x: observed exposure-to-risk in age interval $[x, x + 1)$. E_x is approximated by the midyear population at age x
- l_x: observed number alive at age x
- $m_x = d_x/E_x$: central rate of death at age x
- $q_x = d_x/l_x = \mathbb{P}(x < X < x + 1 | X > x) = \frac{F(x+1)-F(x)}{1-F(x)} = \frac{S(x)-S(x+1)}{S(x)}$: death rate at age x
- ${}_sq_x = \mathbb{P}(x < X < x + s | X > x)$. Note that by definition, $\lim_{s \to 0} \frac{{}_sq_x}{1-{}_sq_x} = h(x)$

2.2 Static advanced-age mortality models

A variety of methodologies have been employed in the past in order to model human mortality. Very early on, Gompertz (1825) suggested that mortality increases exponentially with age during the adult years of life. This is the so-called Gompertz Law, and is observed to hold accurately in the age window from approximately 30 to 80 years of age. Makeham (1860) extended the Gompertz (1825) model by adding an age-dependent component to better capture younger age mortality rates. These models are static and one-dimensional, and cannot be used for mortality forecasting. Later, a number of new approaches were developed using stochastic models such as those of Lee and Carter (1992), Renshaw and Haberman (2006) and Currie, Durban, and Eilers (2004) for mortality forecasting. A useful literature review of mortality models can be seen in Booth and Tickle (2008). However, these models are only suitable for mortality modelling of non-extreme ages, say up to age 90. Due to the sparse, unreliable and volatile nature of advanced age mortality data, there is only a small amount of literature on advanced age mortality modelling and very limited work attempting to model non-extreme and extreme ages in a coherent framework. We summarise some existing static and dynamic models of advanced age mortality in the following subsections.

2.2.1 Heligman-Pollard Model

Heligman and Pollard (1980) modelled the rate of mortality q_x by the following function:

$$\frac{q_x}{1-q_x} = A^{(x+B)^C} + D\exp(-E[\ln x - \ln F]^2) + GH^x, \quad x = 1, 2, 3, \ldots. \tag{1}$$

Here A, B, C, D, E, F, G, H are constants, and the terms $A^{(x+B)^C}$, $D\exp(-E[\ln x - \ln F]^2)$ and GH^x represent early childhood mortality, accidental mortality, and senescent mortality, respectively. GH^x may also be seen as a term capturing the Gompertz-Makeham law of mortality.

However, there is strong empirical evidence that old-age mortality is not Gompertzian (Olshansky and Carnes, 1997). Thus the Heligman-Pollard model suffers from a lack of fit to real data.

Furthermore, $q_x < 1$ for all $x > 0$ in this model. This is a drawback for actuaries as life tables need to be closed in order to value insurance products.

For their mortality models, The Australian Government Actuary (AGA) uses a cubic spline for lower ages (Currie et al., 2004) in conjunction with the Makeham law (asymptotically equivalent to the Heligman-Pollard model) for ages 85 and above (Australian Government Actuary, 2019), and arbitrarily closes the life table at age 110. The Heligman-Pollard model thus warrants careful scrutiny and comparison with competing methods.

2.2.2 Coale-Kisker Method

The Coale-Kisker method (Coale and Guo, 1989; Coale and Kisker, 1990) is another well-known technique for modelling death rates in developed countries (Boleslawski and Tabeau, 2001). The method is recursive, of the form:

$$k(x) = k(x-1) - R, \quad x \geq x_0, \tag{2}$$

where $k(x) = \ln(m_x/m_{x+1})$, R is a constant to be determined, and the extrapolation starts at integer age x_0. Applying the formula up to age $x = x_1$ and using a telescoping sum, we have

$$k(x_0+1) + \cdots + k(x_1) = (x_1 - x_0)k(x_0) - (1 + 2 + \cdots + (x_1 - x_0))R. \tag{3}$$

Solving for R we obtain

$$R = \frac{(x_1 - x_0)k(x_0) + \ln(m_{x_0}) - \ln(m_{x_1})}{1 + 2 + \cdots + (x_1 - x_0)}. \tag{4}$$

The method requires an assumption of the age x_1 at which the life table is closed as well as the value of the central death rate at this closing age. The age from which to start extrapolating also must be subjectively decided. Coale and Kisker (1990) use $x_0 = 84, x_1 = 110, m_{x_1} = 1.0$.

These drawbacks in current classical methods of modeling extreme age mortality are the motivation

behind a desire to develop new methods.

2.3 Extreme Value Theory

One central feature of extreme value theory is that it provides the means to predict extreme values of a variable in a theoretically sound manner, even with limited availability of data for rare events (Embrechts, Kluppelberg, and Mikosch, 2013). The main results in extreme value theory necessary in justifying the Li et al. (2008) approach to the threshold life table are presented below.

Theorem 1 (Fisher and Tippett, 1928). Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed (iid) random variables with cumulative distribution function F, and let $M_n = \max\{X_1, \ldots, X_n\}$, ie. the nth highest order statistic. Suppose there exists a sequence of pairs $\{a_n, b_n\}_{n\in\mathbb{N}}$, where $a_n > 0$, $b_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x), \quad x \in \mathbb{R},$$

where $G(x)$ is a non-degenerate cumulative density function (cdf).

Then $G(x)$ is either *of type*:

Type I (Fréchet):

$$\boldsymbol{\phi}_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases}$$

Type II (Weibull):

$$\boldsymbol{\psi}_\alpha(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

or Type III (Gumbel):

$$\boldsymbol{\Lambda}(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R},$$

where G *of type* H means that for some $a > 0$ and $b \in \mathbb{R}$, $G(x) = H((x-b)/a),\ x \in \mathbb{R}$.

These three types of distributions are special cases of the generalised extreme value distribution (GEV), having cdf

$$H_{\xi;\mu,\sigma}(x) = \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)_+^{-1/\xi}\right\}, \quad x \in \mathbb{R}. \tag{5}$$

Then $\xi > 0$ ($\xi < 0$) corresponds to the Fréchet (Weibull) type distribution, with $\xi = 1/\alpha$ ($\xi = -1/\alpha$), and $\xi = 0$ (understood in the limiting sense as $\xi \to 0$) corresponds to the Gumbel type.

The link to the generalised Pareto Distribution (GPD) used in the threshold life table comes from the following theorem:

Theorem 2 (Pickands, Balkema and de Haan 1974). Suppose $X_1, ..., X_n$ are iid with cdf F, and define $F^{(n)} = \mathbb{P}(M_n - b_n \leq xa_n)$, $x \in \mathbb{R}$. Then $F^{(n)} \to H_\xi$ if and only if there exists some function $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ s.t.

$$\lim_{u \uparrow x_F} \sup_{0<x<x_F-u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0, \tag{6}$$

where $x_F \equiv \sup\{x \in \mathbb{R} : F(x) < 1\} < \infty$,

$$F_u(x) = \mathbb{P}(X - u \leq x | X > u), \tag{7}$$

and $G_{\xi,\beta(u)}(x)$ is the cdf of the generalised Pareto distribution, given by

$$G_{\xi,\beta}(x) = 1 - \left(1+\xi\frac{x}{\beta(u)}\right)_+^{-1/\xi}, \quad x > 0. \tag{8}$$

What this theorem means is that for a large enough threshold u, the conditional excess of X over u given $X > u$ will have a distribution that is approximately generalised Pareto, as long as $F^{(n)}$ has a non-degenerate limiting distribution for some choice of a_n and b_n.

More explicitly, if for large n, $\mathbb{P}(M_n \leq z) \approx H_{\xi;\mu,\sigma}(z)$ for some μ, σ and ξ, then for sufficiently large

u, the distribution function of $X - u$ given $X > u$ is

$$\mathbb{P}(X - u \leq x | X > u) \approx G(x) \tag{9}$$

$$= 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi} \tag{10}$$

on $\{x : 1 + \xi x/\beta > 0\}$ where $\beta = \sigma + \xi(u - \mu)$.

The shape parameter ξ of the generalised Pareto distribution is the corresponding ξ in H_ξ. The GPD becomes a Pareto distribution in the case $\xi > 0$, the exponential distribution when $\xi \to 0$, and a beta distribution if $\xi < 0$ (Embrechts, Resnick, and Samorodnitsky, 1999). When $\xi < 0$, the GPD has a finite upper bound given by $-\beta/\xi$. This enables us to determine the highest attained age ω using the TLT.

Now let $\bar{F}(x) = 1 - F(x)$. Then $\bar{F}(u+x) = \bar{F}(u)\bar{F}_u(x)$ (Balkema and de Haan, 1974), so $F(u+x) = 1 - \bar{F}(u)\bar{F}_u(x)$. Thus we can express $F(x)$ in the form

$$F(x) = \begin{cases} F_1(x), & x \leq u \\ 1 - \bar{F}_1(u)\bar{F}_u(x - u), & x > u \end{cases} \tag{11}$$

where $F_1(x)$ is the assumed cdf below the threshold u. Equation (11) guarantees continuity of F at u, because $\bar{F}_u(0) = 1$ for any u, by (7).

This identity is convenient as we can then model F_1 and F_u separately - the latter of which we have an asymptotic result for in (8). This is exactly the form of model that will be used in the threshold life table.

2.4 Threshold Life Table

The threshold life table (Li et al., 2008) models the age-at-death as a continuous random variable X with cumulative distribution function (cdf) F. The model is piecewise: below a nominated

deterministic threshold age N, $F(x)$ is assumed to follow the Gompertz law of mortality (Gompertz, 1825). Above the threshold age N, $F(x)$ is assumed to follow a GPD, which is justified by the asymptotic result in Theorem 2. The model is completely defined by:

$$F(x) = 1 - \exp\left(-\frac{B}{\ln C}(C^x - 1)\right), \qquad x \leq N, \tag{12}$$

where $B > 0$ and $C > 1$ are constants, and

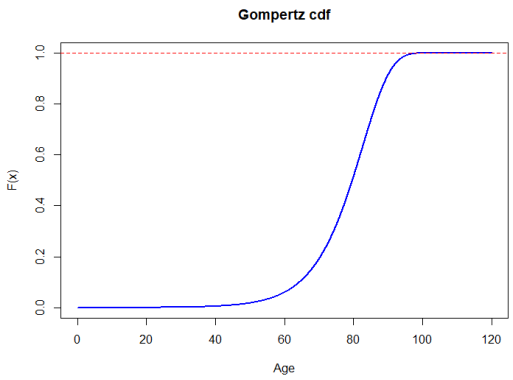
$$F(x) = F_\gamma(x) = \begin{cases} 1 - S(N)(1 + \gamma(\frac{x-N}{\theta}))^{-\frac{1}{\gamma}}, & \gamma > 0,\ x > N \\ 1 - S(N)\exp(-(\frac{x-N}{\theta})), & \gamma = 0,\ x > N \\ 1 - S(N)(1 - |\gamma|(\frac{x-N}{\theta}))^{\frac{1}{|\gamma|}}, & \gamma < 0,\ N < x < N + \frac{\theta}{|\gamma|}, \end{cases} \tag{13}$$

where $\theta > 0$ and $\gamma \in \mathbb{R}$ are constants. This is of the form given in (11), with u replaced by N, $F_1(x)$ identified with $F(x)$ for $x \leq N$ in 12, and 13 corresponding to the component of $F(x)$ for $x > u$ in 11. The cdf $F(x)$ is continuous for $x > 0$, but the corresponding density and hazard rate are not; they are discontinuous at age $x = N$.

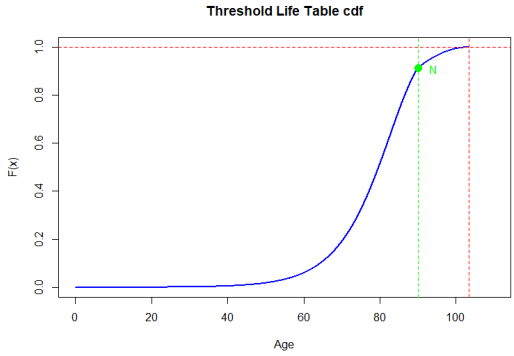
In the Gompertz part of the distribution, (12), B may be interpreted as a 'baseline' force of mortality, ie. the force of mortality of a life at age 0. C controls for the age effect - the amount by which the force of mortality increases with in age.

For the GPD part, (13), θ and γ are scale and shape parameters, and the threshold age N plays the role of a location parameter. Analogous to the Gompertz distribution parameters, θ controls the 'baseline' force of mortality for extreme ages, i.e., the force of mortality of a life at threshold age N, while the parameter γ controls for the age effect for extreme ages. When $\gamma \geq 0$, $F(x)$ has an infinite right endpoint and indefinitely large lifetime ages are potentially possible. But when $\gamma < 0$, $F(x)$ has a finite right end point at $x = N + \theta/|\gamma|$, and ages greater than this cannot occur. Thus the estimated value $\hat{\gamma}$ of γ from a dataset is relevant to the question of whether there is a finite upper limit to the lifetime of an individual from the population ($\hat{\gamma} < 0$) or no such upper limit ($\hat{\gamma} > 0$). A stylized comparison between a generic threshold life table fit with $\gamma < 0$ and a

generic Gompertz model fit is shown in Figure 1:



(a) A Gompertz model, with $F(x) \to 1$ as $x \to \infty$, but $F(x) < 1$ for all x.



(b) A threshold life table model, where $F(x) = 1$ at some finite x, with a discontinuity in the density at age N.

Figure 1: *Difference between the Gompertz and Threshold Life Table cumulative density functions*

The threshold life table is more appropriately applied to cohorts than to period data because observations are assumed to be independently and identically distributed with constant parameters with respect to time - so the continuous age-at-death random variable should have the same distribution for all individuals to which the method is applied. Thus, our analysis in this thesis will be cohort-by-cohort.

2.5 Predictive Models

2.5.1 Combined Models

Models in which the lifetime distribution describes a fixed birth-year or death-year cohort do not lend themselves naturally to prediction. Some existing models forecast future advanced age mortality by combining a static model with an external stochastic mortality model for non-extreme ages. The Lee-Carter model (Lee and Carter, 1992) is an important example. In these types of models, the static advanced age mortality component can then be fitted to the predicted death rates of these birth-year or death-year cohorts. For example, Lee and Carter (1992) combined

the component of their model for ages below 85 with the Coale-Kisker method. This combined model however overlooks the inherent relationship between non-extreme and extreme ages, as well as the temporal dependence of extreme-age mortality rates. Later, Li et al. (2008) combined the Lee-Carter model with their threshold life table method. In this framework, the mortality of non-extreme ages is modelled twice, first using the Lee-Carter model and then using the TLT model. But this redundancy makes the model difficult to interpret. The temporal dependence of extreme-age mortality rates is also ignored in this setup.

2.5.2 Watts-Dupuis-Jones Model

An extreme value approach to advanced age mortality prediction is taken by Watts, Dupuis, and Jones (2006). They fitted a generalised extreme value (GEV) distribution (5) to Japanese mortality data using the 10 highest ages at death recorded annually between 1980 and 2000.

To incorporate time variability, they modelled the parameters μ and σ as a function of time:

$$\mu(t) = \mu_0 + \mu_1 t^*, \text{ and } \sigma(t) = \exp(\sigma_0 + \sigma_t t^*),$$

where $t^* = \frac{t-1980}{20}$ is the standardized year of death in their data, for $0 \leq t^* \leq 1$. In assuming this parametric form, Watts et al. (2006) force the upper limit to the human lifespan, ω, to be a linear function of time. In this thesis, we allow time dependence of the parameters, but we do not impose such linear constraints. Note that the Watts-Dupuis-Jones model can only be used to forecast the highest attained age; the dynamic model we propose in Section 5 can be used to forecast both future mortality rates and highest attained ages.

2.5.3 Cairns-Blake-Dowd Model

Other methods in the literature used to predict extreme age mortality include the Cairns-Blake-Dowd (CBD) model (Cairns, Blake, Dowd, Coughlan, Epstein, and Khalaf-Allah, 2011). The model

is defined by

$$\log h_{i,j} = \kappa_{0,j} + \kappa_{1,j}(x_i - \bar{x}), \tag{14}$$

where $h_{i,j}$ is the force of mortality in year j at age x_i, and $\kappa_{0,j}, \kappa_{1,j}$ are latent time-varying variables typically modelled as a bivariate random walk with drift. The model can be rewritten in the form of a Gompertz model with time-varying parameters:

$$\log h_{i,j} = \log B_j + x_i \log C_j, \tag{15}$$

where $B_j = \exp(\kappa_{0,j} - \kappa_{1,j}\bar{x})$ and $C_j = \exp(\kappa_{1,j})$.

Extensions to the CBD model include adding a quadratic term (Currie, 2011):

$$\log h_{i,j} = \kappa_{0,j} + \kappa_{1,j}(x_i - \bar{x}) + \kappa_{2,j}[(x_i - \bar{x})^2 - \sum (x_i - \bar{x})^2 / n_a], \tag{16}$$

where n_a is the number of ages used to fit the model.

The CBD model is widely used in practice and in Section 5 we use it to benchmark the effectiveness of our proposed dynamic smooth threshold life table (DSTLT) method.

3 Data and Methodology

3.1 Data Source

This thesis combines Netherlands population data from the Human Mortality Database (HMD) and a dataset provided by Centraal Bureau voor de Statistiek (CBS). We will refer to data from these sources as HMD data and CBS data, respectively.

The Human Mortality Database is a joint initiative of the Department of Demographics at the Uni-

versity of California, Berkeley and the Max Planck Institute for Demographic Research. It provides detailed mortality and population data for 40 countries, including the Netherlands. The HMD data for the Netherlands is gathered from various sources, including the Netherlands Interdisciplinary Demographic Institute (NIDI), Eurostat, and the CBS (Jasilionis, 2018). The data contains age and gender-specific death and number-alive counts from 1850-2016. The counts are given annually and thus are interval censored; they are also right-censored at age 100.

The CBS data consists of the ages at death (to the nearest day) of all age ≥ 92 Dutch residents whose year of death was between 1986-2015. The data is not in any way censored. Calculating the number of persons alive using the dataset requires two assumptions. One is that every individual in the cohort concerned has died (and thus occurs as an observation) by the end of our observation period, since the dataset only records deaths and not numbers alive. The other is that net migration is negligible for high (>92) ages. With these caveats, we can deduce the number of persons alive at each age and year - so approximation techniques for the exposure such as that employed in the HMD data (Jasilionis, 2018) need not be used.

Since the two data sets come from quite different sources, there are possible discrepancies between them. The CBS data set contains only death counts, so to check whether the two datasets do indeed describe the same population, we computed the differences in the number of gender, age and year-specific deaths between the two datasets. These are shown in Table 13 in Appendix A. For the HMD data, the death count for each cell is given in the raw data. For the CBS data, we calculate the age-at-death in years by dividing the age-at-death in days (as given in the data) by 365.25. The number of deaths in each cell is then found by counting the number of deaths that fall in the corresponding age and year interval.

Apart from two discrepancies in a year×age-group cell of 17 and 11 persons, the remainder are in the single digits[1]. This is a small difference given the large number of deaths in each year×age-group cell, which are typically in the thousands or several hundreds (see Table 14 in Appendix A).

For our analysis, we ignore these minor discrepancies and treat the two datasets as compatible to

[1]These discrepancies necessarily occur at ages less than 100

be combined together.

3.2 Data Combination and Generation

The TLT models the lifetime distribution for a given population, and does not take into account the exposure-to-risk. Consequently, raw death counts cannot be directly used to fit the model, because migration may result in changing exposures-to-risk. For example, positive net migration would increase the exposure-to-risk and thus the total number of deaths. A TLT model fitted to the raw death counts would then overestimate death rates, since there are more deaths than would otherwise be observed if the exposure-to-risk was constant.

To overcome this problem, we fit the model using a constructed hypothetical cohort, with death rates calculated from the raw data $q_x = d_x/l_x$. The steps as are follows:

1. Set $l'_{65} = l_{65}$, where l'_{65} is the number of persons alive at age 65 in our hypothetical cohort.
2. For $x = 65, 66, \ldots, \tau$, where τ is the observed maximum attained age for the cohort in question:
   (a) Calculate the number of deaths at age x in our hypothetical cohort from $d'_x = l'_x q_x$.
   (b) Calculate $l'_{x+1} = l'_x - d'_x$.

The combined dataset is constructed as follows. For ages above 92, the death counts in each age interval are computed using the CBS data. The number of persons alive at age x ($x > 92$) is then calculated as $l_x = \sum_{i=x}^{\tau} d_i$, where τ is the largest observed age at death in the raw data. We can then compute the probability of death at age x, $q_x = d_x/l_x$. This assumes all changes in population numbers are from deaths. This construction is consistent with the extinct cohorts method (Vincent, 1951) for exposure calculation. For ages below and equal to 92, we extract q_x from the cohort life table of the HMD. The q_x are computed directly from the raw data and no smoothing at older ages are performed (Human Mortality Database, 2019). Combining the computed probabilities of death q_x obtained from the two databases gives an augmented dataset of the Netherlands for all birth

cohorts from 1893 to 1908[2]. We plot the observed death probabilities, q_x, in Figure 2. Since there are zero death probabilities for some very high ages (see Figure 2), we right-censor in the combined dataset at age 108 for females and at age 107 for males for modelling purposes.

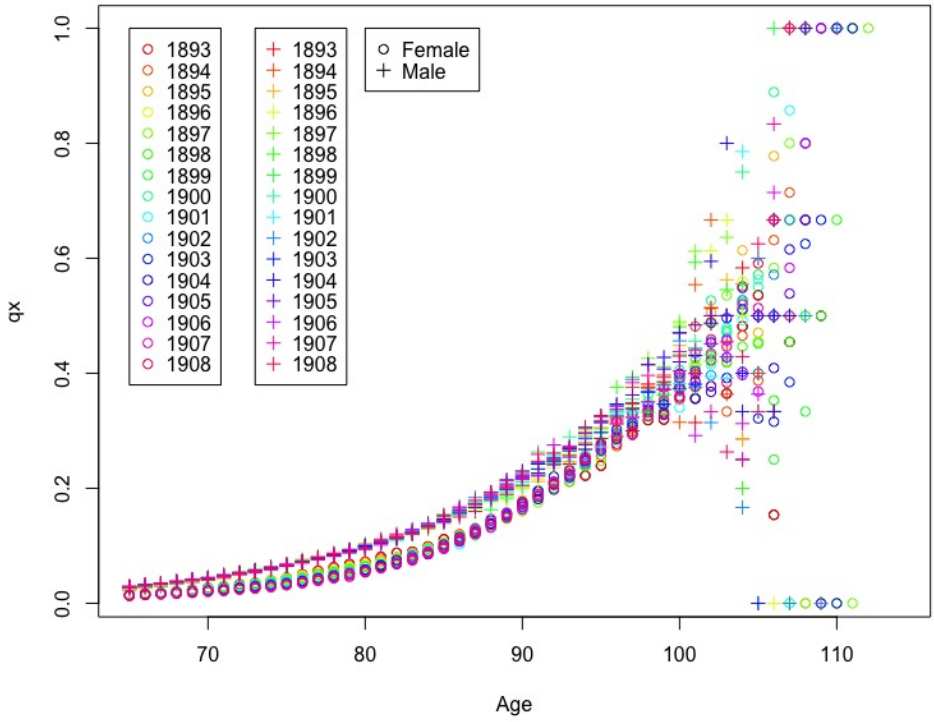


Figure 2: *Observed Netherlands Death Probabilities*

3.3 Methodology Overview

The methodology in this thesis is comprised of two main parts. In the first - static - part, we introduce a smoothing constraint to the threshold life table and compare the smooth threshold life table (STLT) with the original threshold life table (TLT), the Heligman-Pollard model, and the Coale-Kisker method. This is done for the 9 different cohorts from the years 1893-1901 respectively. Figure 3 illustrates how we proceed, for the female cohort born in 1901, using HMD data only.

[2]We use the data of cohorts 1893 to 1901 for in-sample fitting, and cohorts 1902 to 1908 for out-of-sample forecasting

In the second part, we make the smooth threshold life table dynamic by modelling parameters as a function of time. This new model is then used for prediction and compared with predictions from the Cairns-Blake-Dowd model. We consider mortality modelling and forecasting for ages 65 and above only, as those ages are most relevant for retirement-related insurance products and longevity management.

For the static part, our main steps are as follows:

1. For each annual cohort born between 1893-1901, use the CBS data to calculate number alive at each integer age above 92 by summing all future death counts for the cohort. Calculate death rates at each age above 92 by dividing death counts by number alive at that age.

2. Generate death counts using HMD data alone.

3. Combine the CBS and HMD data, and generate death counts from the combined data set.

4. For each cohort (slanting green lines in Figure 3), fit a Gompertz model using data before the threshold age N (blue line). For ages >N (black solid line), fit a generalised Pareto distribution. This is first done using HMD data alone, and then repeated on the combined dataset. When HMD data alone is used, only data up to age 100 are available (as shown in Figure 3). When the augmented dataset is used, the data is not censored and all ages at death data are available. However, we manually interval and age-censor the data in our study for modelling purposes.

5. Given the parameter estimates of the Generalised Pareto distribution obtained in Step 4, model the tail behaviour of cohort death rates (black dotted line), and take the upper bound (possibly infinite) of the fitted distribution as the estimated highest attainable age.

6. Compare the fitted STLT with the TLT, Heligman-Pollard model and Coale-Kisker methods.

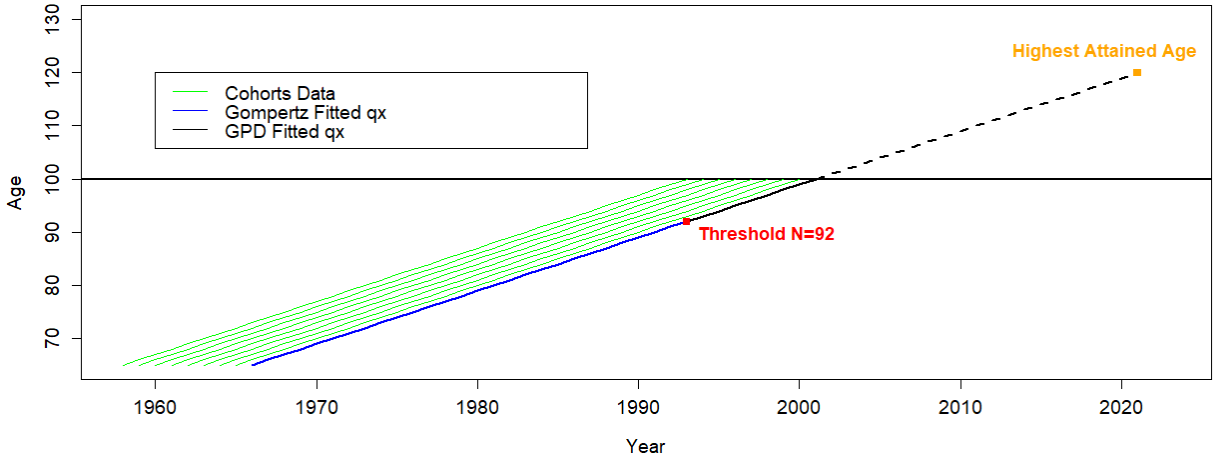


Figure 3: *Schematic for threshold life table fit to the cohort born in 1901, N=92*

The methodology and model specification for the dynamic part is discussed in Section 5.

3.4 Estimation Methods

3.4.1 TLT Model

For the TLT model as set up in (12) and (13), the likelihood for censored data (single-year interval censored from age 65 to age $\tau - 1$, and right censored at age τ, where τ is the observed maximum attained age for the cohort) is given by

$$L(B, C, \gamma; N) = \left[\prod_{x=65}^{N-1} \left(\frac{S(x) - S(x+1)}{S(65)} \right)^{d_x} \prod_{x=N}^{\tau-1} \left(\frac{S(x) - S(x+1)}{S(65)} \right)^{d_x} \right] \left(\frac{S(\tau)}{S(65)} \right)^{l_\tau}. \qquad (17)$$

With careful algebraic manipulation, the log-likelihood can be additively decomposed into two parts, the first being

$$l_1(B,C;N) = \sum_{x=65}^{N-1} \ln\left(\frac{S(x)-S(x+1)}{S(65)}\right)^{d_x} + l_N((\ln S(N) - \ln S(65))$$
$$= \sum_{x=65}^{N-1} d_x \ln(S(x)-S(x+1)) + l_N \ln(S(N)) - l_{65}\ln(S(65)), \tag{18}$$

(because $\sum_{x=65}^{N-1} d_x + l_N = l_{65}$) where $S(x) = \exp(-B/\ln C(C^x - 1))$, $65 \le x \le N$, and the second being

$$l_2(\gamma,\theta;N) = \sum_{x=N}^{\tau-1} \ln\left(\frac{S(x)-S(x+1)}{S(N)}\right)^{d_x} + l_\tau((\ln S(\tau) - \ln S(N))$$
$$= \sum_{x=N}^{\tau-1} d_x \ln\left(S(x)-S(x+1)\right) + l_\tau \ln\left(S(\tau)\right) - l_N \ln(S(N)), \tag{19}$$

because $\sum_{x=65}^{\tau-1} d_x + l_\tau = l_N$ where $S(x) = S(N)(1+\gamma((x-N)/\theta))^{-1/\gamma}$, $N \le x < \tau$.

This shows that the parameters for the Gompertz and GPD parts can be estimated separately, leading to the following fitting algorithm:

1. For a given integer N (we choose $N = 85$),
   (a) Find $(\hat{B},\hat{C})^\top = \arg\max_{B,C} l_1$
   (b) Find $(\hat{\gamma},\hat{\theta})^\top = \arg\max_{\gamma,\theta} l_2$
   (c) Compute the value of the profile log-likelihood, $l_p = l_1(\hat{B},\hat{C};N) + l_2(\hat{\gamma},\hat{\theta};N)$
2. Repeat step 1 for each $N = 85, 86, \ldots, U$, where U is the upper limit of the threshold age N
3. Find $\hat{N} = \arg\max_N l_p$

Empirically, we find that the log-likelihood levels off after around age 95 for all cohorts. We thus set U as the minimum age at which the level-off occurs so as to retain the maximum number of observations available for fitting the extreme value part of the model. This resulted in setting U as age 97 and age 98 for HMD only and HMD+CBS data, respectively.

3.4.2 Smoothed TLT model

For the smooth-constrained model, we set $h_1(N) = h_2(N)$, where h_1 and h_2 are the hazard functions of the Gompertz and GPD parts of the distribution respectively. This ensures a smooth change over in hazards between the two parts, and allows us to eliminate one parameter, say θ. This leads to the following parameter constraint:

$$\theta = \frac{1}{BC^N} \tag{20}$$

(see Appendix B for derivation). For the STLT, the two component distributions are therefore

$$F(x) = 1 - \exp\left(-\frac{B}{\ln C}(C^x - 1)\right) \qquad x \leq N; \tag{21}$$

and

$$F(x) = F_\gamma(x) = \begin{cases} 1 - S(N)(1 + \gamma(C^N B(x-N)))^{-\frac{1}{\gamma}}, & \gamma > 0,\ x > N \\ 1 - S(N)\exp(-(C^N B(x-N))), & \gamma = 0,\ x > N \\ 1 - S(N)(1 - |\gamma|(C^N B(x-N)))^{\frac{1}{|\gamma|}}, & \gamma < 0,\ N < x < N + \frac{\theta}{|\gamma|}, \end{cases} \tag{22}$$

The log-likelihood no longer separates into two disjoint components, and thus we must fit (17) directly to the entire data set, after taking into account (20). This means that we must find the maximum likelihood estimates of B, C, and γ jointly.

It turns out that the maximum likelihood estimates (MLEs) for B and C in the Gompertz distribution are often very close to their boundary values ($B \approx 0$, $C \approx 1$), and numerical algorithms for computing them tend to be unstable (Chartres and Stepleman, 1972). Consequently we reparameterise B as e^α and C as e^{e^δ}. The parameter space of both α and δ is $(-\infty, \infty)$, improving numerical stability. Having found the MLEs $\hat{\alpha}$, $\hat{\delta}$ we have that $\hat{B} = e^{\hat{\alpha}}$ and $\hat{C} = e^{e^{\hat{\delta}}}$ by the invariance property of MLEs (Tan and Drossos, 1975). This reparameterisation must be taken into

account when computing the approximate variance of the estimated highest attained age $\hat{\omega}$ via the delta method, as described in the next section.

3.5 Highest Attained Age

Let $Y = X - N$ denote the exceedance of an individual lifetime X over threshold age N. In both the TLT and STLT, we model the distribution of Y conditional on $X > N$ as a GPD of the form:

$$F_Y(y) = \begin{cases} 1 - \left(1 + \gamma \frac{y}{\theta}\right)^{-1/\gamma}, & y > 0,\ \gamma > 0 \\ 1 - \left(1 + \gamma \frac{y}{\theta}\right)^{-1/\gamma}, & 0 < y \leq -\theta/\gamma,\ \gamma < 0 \\ 1 - e^{-y/\theta}, & y > 0,\ \gamma = 0. \end{cases} \tag{23}$$

Now let $M_n = \max\{Y_i,\ i = 1, 2, 3, ..., n\}$. If $\gamma < 0$, then

$$F_{M_n}(y) = \begin{cases} 0, & y < 0 \\ (F_Y(y))^n, & 0 < y < -\theta/\gamma \\ 1, & y \geq -\theta/\gamma, \end{cases}$$

and so

$$\lim_{n \to \infty} F_{M_n}(y) = \begin{cases} 0, & y < -\theta/\gamma \\ 1, & y \geq -\theta/\gamma. \end{cases}$$

This implies that under this model, as population size n at the threshold age N becomes large, the highest age at death of individuals in the population converges to

$$\omega = N - \theta/\gamma = N + \frac{\theta}{|\gamma|}, \tag{24}$$

the theoretical end point of the threshold life table.

The asymptotic variance of $\hat{\omega}$ can be computed via the Delta Method (Oehlert, 1992), as follows.

Suppose we are given a vector of parameters $\boldsymbol{\beta}$, a corresponding estimator $\hat{\boldsymbol{\beta}}$, and a multivariate function $h : \mathbb{R}^n \to \mathbb{R}$. Then $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \xrightarrow{D} N(\mathbf{0}, \Sigma) \Rightarrow \sqrt{n}(h(\hat{\boldsymbol{\beta}})-h(\boldsymbol{\beta})) \xrightarrow{D} N(0, \nabla h(\boldsymbol{\beta})^T \cdot (\Sigma) \cdot \nabla h(\boldsymbol{\beta}))$, where ∇ denotes the vector differential operator.

Applying this to $\hat{\omega}$, we get

$$\operatorname{Var}(\hat{\omega}) = \begin{pmatrix} \frac{\partial \omega}{\partial \gamma} & \frac{\partial \omega}{\partial \theta} \end{pmatrix} I(\gamma, \theta)^{-1} \begin{pmatrix} \frac{\partial \omega}{\partial \gamma} \\ \frac{\partial \omega}{\partial \theta} \end{pmatrix} \tag{25}$$

$$= \begin{pmatrix} \frac{\theta}{\gamma^2} & -\frac{1}{\gamma} \end{pmatrix} I(\gamma, \theta)^{-1} \begin{pmatrix} \frac{\theta}{\gamma^2} \\ -\frac{1}{\gamma} \end{pmatrix}, \tag{26}$$

where

$$I(\gamma, \theta) = -\mathbb{E}[\mathrm{H}(\gamma, \theta)] = -\mathbb{E}\begin{bmatrix} \frac{\partial^2 l}{\partial \gamma^2} & \frac{\partial^2 l}{\partial \gamma \partial \theta} \\ \frac{\partial^2 l}{\partial \theta \partial \gamma} & \frac{\partial^2 l}{\partial \theta^2} \end{bmatrix}.$$

Relying on the asymptotic normality of maximum likelihood estimates (Young, 2019), an approximate 95% confidence interval for ω can be written as

$$(\hat{\omega} - 1.96\sqrt{\operatorname{Var}(\hat{\omega})}, \hat{\omega} + 1.96\sqrt{\operatorname{Var}(\hat{\omega})}). \tag{27}$$

For the constrained model, $\theta = \frac{1}{BC^N}$, so

$$\omega = N - (\gamma C^N B)^{-1} = N - (\gamma (e^{e^{\delta}})^N e^{\alpha})^{-1} \tag{28}$$

under our reparameterisation. This gives

$$
\begin{aligned}
\mathrm{Var}(\hat{\omega}) &= \begin{pmatrix} \frac{\partial \omega}{\partial \alpha} & \frac{\partial \omega}{\partial \delta} & \frac{\partial \omega}{\partial \gamma} \end{pmatrix} I(\alpha, \delta, \gamma)^{-1} \begin{pmatrix} \frac{\partial \omega}{\partial \alpha} \\ \frac{\partial \omega}{\partial \delta} \\ \frac{\partial \omega}{\partial \gamma} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\gamma (e^{e^{\delta}})^{N} e^{\alpha}} & \frac{N (e^{e^{\delta}})^{-N} e^{\delta}}{\gamma e^{\alpha}} & \frac{1}{\gamma^2 (e^{e^{\delta}})^{N} e^{\alpha}} \end{pmatrix} I(\alpha, \delta, \gamma)^{-1} \begin{pmatrix} \frac{1}{\gamma (e^{e^{\delta}})^{N} e^{\alpha}} \\ \frac{N (e^{e^{\delta}})^{-N} e^{\delta}}{\gamma e^{\alpha}} \\ \frac{1}{\gamma^2 (e^{e^{\delta}})^{N} e^{\alpha}} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\gamma C^N B} & \frac{N \ln C}{\gamma C^N B} & \frac{1}{\gamma^2 C^N B} \end{pmatrix} I(\alpha, \delta, \gamma)^{-1} \begin{pmatrix} \frac{1}{\gamma C^N B} \\ \frac{N \ln C}{\gamma C^N B} \\ \frac{1}{\gamma^2 C^N B} \end{pmatrix} \\
&= \frac{1}{(\gamma C^N B)^2} \left[\begin{pmatrix} 1 & N \ln C & \gamma^{-1} \end{pmatrix} I(\alpha, \delta, \gamma)^{-1} \begin{pmatrix} 1 \\ N \ln C \\ \gamma^{-1} \end{pmatrix} \right],
\end{aligned}
$$

where

$$
I(\alpha, \delta, \gamma) = -\mathbb{E} \begin{bmatrix} \frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \delta} & \frac{\partial^2 l}{\partial \alpha \partial \gamma} \\ \frac{\partial^2 l}{\partial \alpha \partial \delta} & \frac{\partial^2 l}{\partial \delta^2} & \frac{\partial^2 l}{\partial \delta \partial \gamma} \\ \frac{\partial^2 l}{\partial \gamma \partial \alpha} & \frac{\partial^2 l}{\partial \gamma \partial \delta} & \frac{\partial^2 l}{\partial \gamma^2} \end{bmatrix}. \tag{29}
$$

Here we note that $I(\hat{\alpha}, \hat{\delta}, \hat{\gamma}) \to I(\alpha, \delta, \gamma)$ and $(\hat{\alpha}, \hat{\delta}, \hat{\gamma}) \to (\alpha, \delta, \gamma)$ by the asymptotic unbiasedness and consistency of maximum likelihood estimates (Young, 2019), allowing us to compute the approximate variance of $\hat{\omega}$.

4 Results for the STLT

Our results for the STLT are presented in two main subsections. One focuses on the difference in fits between our proposed smooth threshold life table (STLT) and existing models. The other focuses on the implications of these differences for an annuity valuation. To illustrate the methodology, our presented results are for a representative cohort - namely, Netherlands females born in 1901. Further calculations confirm that the resulting implications and conclusions remain largely consistent across all cohorts in our dataset.

4.1 Comparison of Methods

Our first aim is to investigate the relative merits of our smoothing constraint on the threshold life table. We also highlight the importance of having reliable data at high ages. We then compare the smooth threshold life table with two other models commonly used in practice: the Heligman-Pollard model and the Coale-Kisker method.

4.1.1 Effect of Smoothing Constraint and Combining Data

In the threshold life table method the estimate of the γ parameter, $\hat{\gamma}$, is often poorly determined. See for example the 3D likelihood surface plot in Figure 4, produced for the female 1901 cohort using only HMD data:

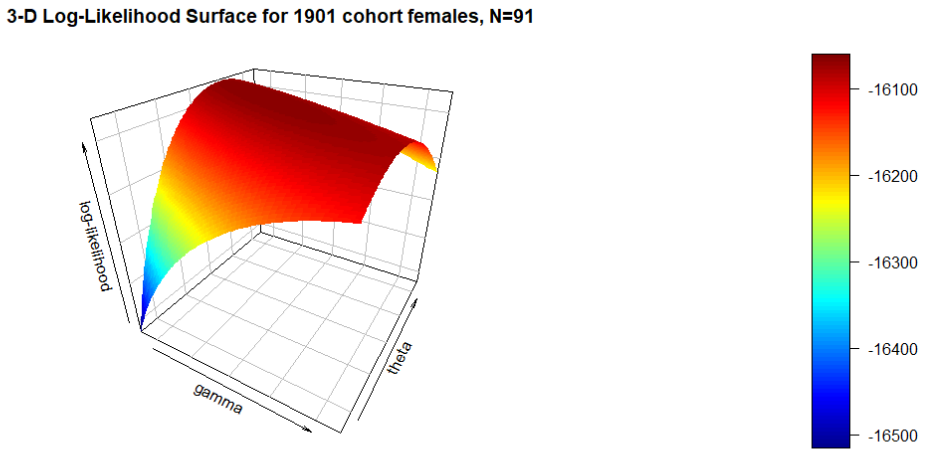


Figure 4: *3-D Log-Likelihood Surface plot for Females born in 1901, N=91; HMD data*

This plot shows that the curvature of the log-likelihood is noticeable in the θ direction and this results in a well-determined θ. However, the log-likelihood surface is flat in the γ direction, resulting in a high standard error for $\hat{\gamma}$. For both the TLT and STLT models, Table 1 shows the estimates for γ, ω with their standard errors, and the estimated threshold age N.

Table 1: Estimated γ, highest attained age ω, and threshold age N, female 1901 cohort

Regime	$\hat{\gamma}$	SE($\hat{\gamma}$)	$\hat{\omega}$	SE($\hat{\omega}$)	N
TLT, HMD only	-0.307	0.0143	106.33	0.526	91
STLT, HMD only	-0.237	0.0451	108.98	2.245	97
TLT, HMD+CBS	-0.240	0.0070	109.59	0.388	91
STLT, HMD+CBS	-0.191	0.0132	111.78	0.976	97

The estimated highest attained ages for the HMD-only regimes are smaller than those for the regimes incorporating the CBS data. For the TLT model using the HMD-only data, the upper bound of the 95% confidence interval for ω is less than 108. As the highest-aged survivor in this cohort lived to 108 years, this casts some doubt on the suitability of using the TLT model and HMD data alone. Incorporating the CBS with the HMD data significantly reduces the standard errors of $\hat{\gamma}$ and $\hat{\omega}$. In this case, the standard errors under the STLT model are higher due to a higher estimated threshold age N, which leads to less data being used to fit the GPD.

Figures 5 and 6 show the female 1901 cohort data and the fitted death rate lines for the TLT and STLT models in the HMD and HMD+CBS datasets.

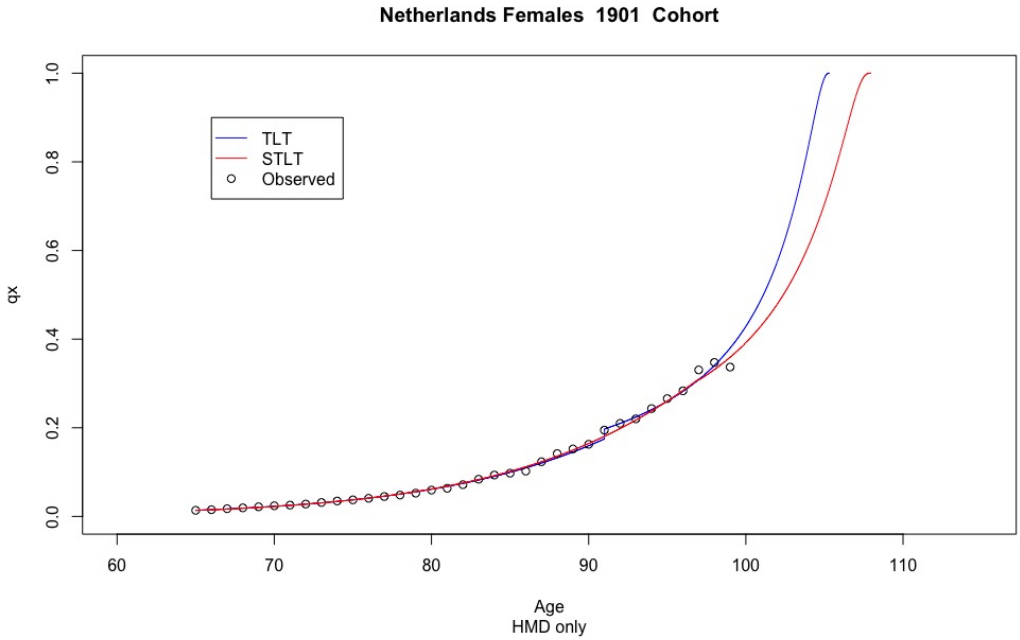


Figure 5: *TLT and STLT fitted curves with HMD data only, female 1901 cohort*

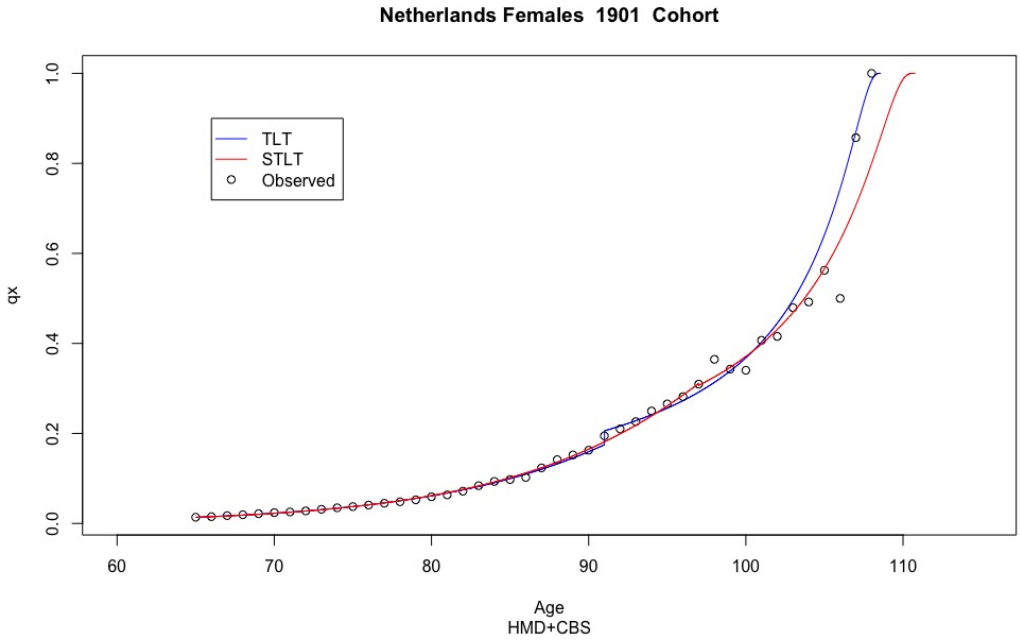


Figure 6: *TLT and STLT fitted curves with HMD and CBS data, female 1901 cohort*

With the TLT fit, there is a discontinuity in the fitted death rate function at the threshold age, due to the piecewise specification of the probability density functions. The smoothed threshold life table removes this discontinuity. The resulting smooth transition altogether is in accordance with our intuition that there should not be a sudden change in mortality experience at any age, including at the threshold. The fitted TLT and STLT models using HMD and augmented datasets for females and males and all cohorts 1893-1901 are presented in Figures 18 - 21 in Appendix C.

Finally, we examine the residual plots for each model:

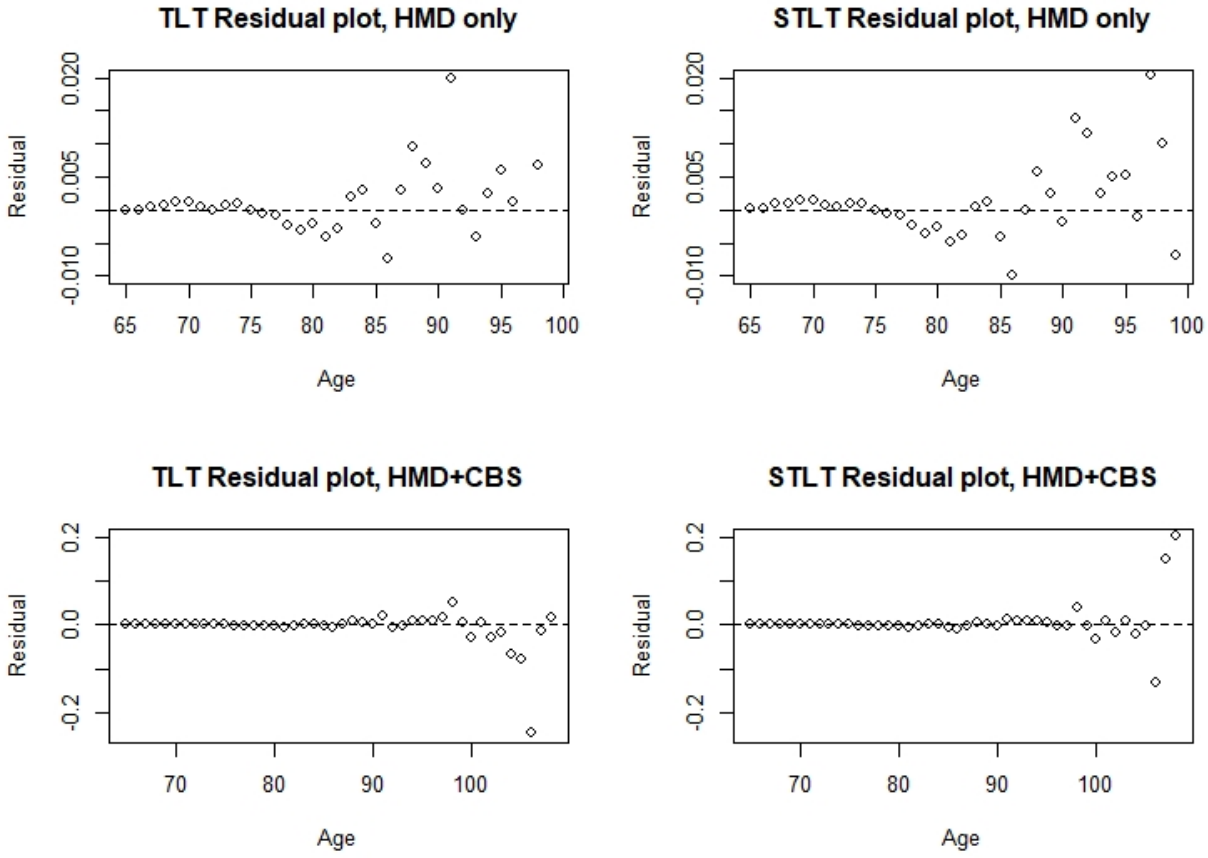


Figure 7: *Residuals under each regime, female 1901 cohort*

The plots show that adding the smoothing constraint does not substantially reduce the goodness of fit of the model for this cohort. To quantify the differences between the STLT and TLT, we examine the sum of squared errors (SSE) for both models, defined by

$$\text{SSE} = \sum_{x=65}^{\tau} (q_x - \hat{q}_x)^2, \tag{30}$$

where q_x and $\hat{q}_x$ are the observed and estimated death rates respectively at age x. The SSEs for the TLT and STLT models for both females and males are in Table 2.

Table 2: SSEs for the TLT and STLT Models, HMD+CBS data

	Females		Males	
Cohort	TLT	STLT	TLT	STLT
1893	0.17	0.18	0.45	0.45
1894	0.20	0.16	0.48	0.50
1895	1.00	0.93	0.26	0.26
1896	0.16	0.10	0.61	0.69
1897	0.72	0.80	0.34	0.39
1898	0.10	0.08	0.35	0.28
1899	0.33	0.27	0.12	0.14
1900	0.84	0.74	0.54	0.50
1901	0.08	0.08	0.51	0.51

The STLT provides no bigger SSEs than the TLT for all but six of the 18 cohorts, and for those six the difference is small. For male cohort 1897, $\hat{\gamma}$ is zero under the TLT when using the HMD data only, resulting in a constant q_x past the threshold age - see Figure 19. After including the CBS data, $\hat{\gamma}$ becomes negative for the same cohort (see Figure 21). For the STLT fits however, we always obtain a negative $\hat{\gamma}$. We note that the smoothing constraint in the STLT constitutes a functional link between the parameters of the Gompertz and GPD distributions, which tends to make the extreme age modelling of STLT more robust than the TLT model; see Figures 18 - 21 in Appendix C.

Overall, strong improvements in interpretability for both the TLT and STLT are seen when incorporating the CBS data. At this stage, we conclude that adding the smoothing constraint improves on the threshold life table as an overall description of the data.

4.1.2 Smooth Threshold Life Table vs Current Methodology

The Heligman-Pollard model is used by the Australian Government Actuary (AGA) to model mortality for ages 85 and above (Australian Government Actuary, 2019), along with a smoothing spline (Currie et al., 2004) approach for lower ages. The model thus warrants careful scrutiny and comparison with our proposed smooth threshold life table. The Coale-Kisker method is used widely in the insurance industry due to its ease of application and lends itself as another good benchmark to compare our model to.

Following the methodology employed by the AGA, we fitted the Heligman-Pollard model using death rates from age 85 and above for the 1901 cohort of females. Meanwhile, for the Coale-Kisker method we started the extrapolation at age 85 and chose the end point of the extrapolation as age 108 (the highest observed attained age for this cohort). That is, we set $m_{108} = 1$. The fits for the these two models, along with the STLT fit, are displayed in Figure 8.

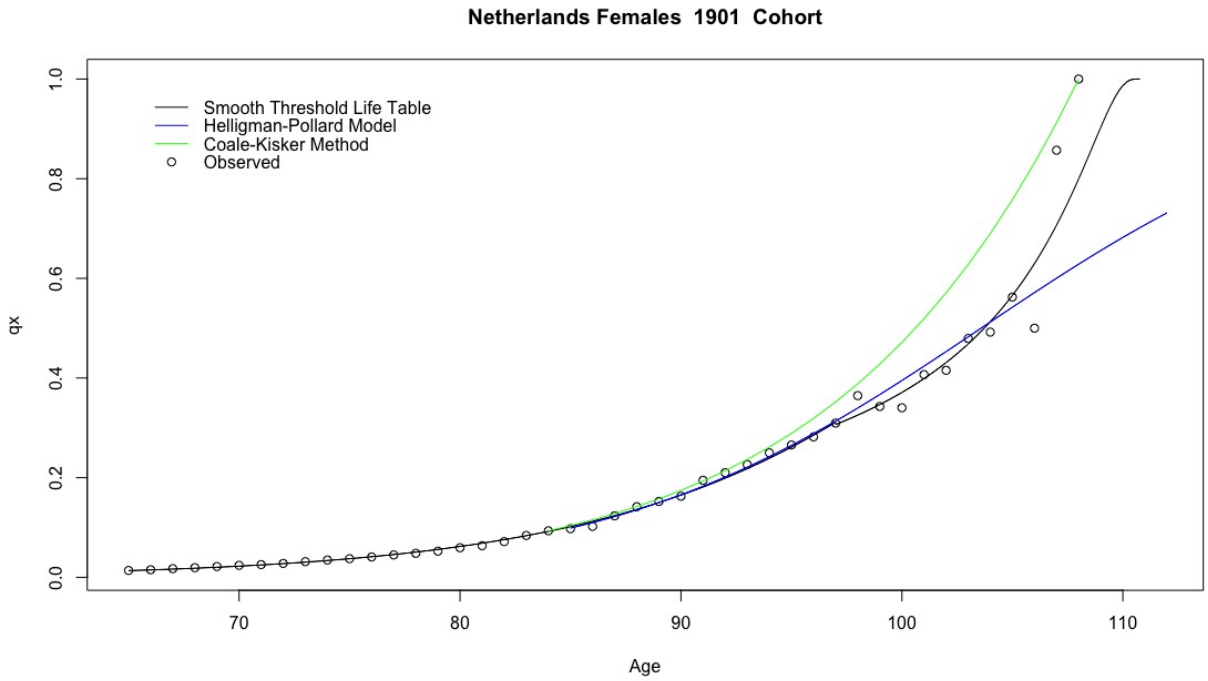


Figure 8: *Fitted lines for STLT, Heligman-Pollard and Coale-Kisker methods, female 1901 cohort*

Here we can make several observations. Firstly, the curve fit with the Coale-Kisker method consistently overestimates death rates past age 92. Inaccuracy at extreme ages is likely with this method as the fitted curve is depends strongly on the start and end points of the extrapolation - an obvious weakness of the technique. Mathematically, we have that R as defined in (4) satisfies (3), under the model. This means that the predicted $\ln(m_{x_1})$ equals our assumed value of $\ln(m_{x_1})$ necessarily. Further, the model implies that

$$\begin{aligned} \ln m_x - \ln m_{x+1} &= \ln m_{x-1} - \ln m_x - R \\ \Rightarrow \ln m_{x+1} &= 2\ln m_x - \ln m_{x-1} + R, \quad x \geq x_0. \end{aligned}$$

Empirically, this imposed form does not fit the data well.

The Heligman-Pollard model by contrast fits the data reasonably well up until around age 105 but severely underestimates the death rate at higher ages. The Heligman-Pollard model is essentially a Gompertz law at high ages. This can be seen from the formulation of the Heligman-Pollard model (1). As x becomes large, the first term tends to 0 because $A < 1$. The second term also tends to 0 since $E > 0$. This leaves

$$\begin{aligned} \frac{q_x}{1-q_x} &\approx GH^x \\ \Rightarrow \ln\left(\frac{q_x}{1-q_x}\right) &\approx \ln(G) + x\ln(H). \end{aligned}$$

Now $\frac{q_x}{1-q_x}$ can be seen as the discrete version of h_x, since $\lim_{s\to 0}\frac{{}_sq_x}{1-{}_sq_x} = h_x$. Thus the Helligman-Pollard model can be seen as asymptotically equivalent to the Gompertz model and assumes that the force of mortality increases exponentially with age. In assuming an exponential extrapolation, the Heligman-Pollard model is clearly inappropriate empirically.

The SSE for the STLT is significantly smaller than for the Heligman-Pollard model and Coale Kisker method[3]. Other cohorts of both genders show similar fitted patterns with the female 1901 cohort, as detailed in Figures 22 and 23 in Appendix D.

[3]SSE(STLT)=0.0836, SSE(Heligman-Pollard)=0.216, and SSE(Coale-Kisker)=0.278.

4.2 Application

To investigate the practical impact of the smooth threshold life table, we value a theoretical life annuity starting at age 85 that pays 1 dollar in arrears to living policyholders. This is again done using the 1901 female cohort. An annual effective real interest rate of 5% is assumed. Through a simulation approach, we compute the mean, value at risk (VaR), and conditional tail expectation (CTE) of the present value (PV). For the VaRs and CTEs, we examine the 95%, 99.5%, 99.9%, and 99.99% confidence levels. The 95% and 99.99% levels are chosen to emphasize the differences between the STLT and Heligman-Pollard model, while the 99.5% and 99.9% levels are chosen to comply with the Solvency II (Institute and Faculty of Actuaries, 2016) and Basel III (Basel Committee on Banking Supervision, 2016) standards respectively. The evaluation is done using fitted death rates under the STLT, Heligman-Pollard, and Coale-Kisker methods separately. The STLT is fitted with data from ages 65-108, and the Heligman-Pollard and Coale-Kisker methods are fitted with data from ages 85 onwards. The endpoint for the Coale-Kisker extrapolation is set at 108, as before. We also value the annuity using empirical death rates, by setting $\hat{q}_x = q_x$ in (34).

The risk measures are chosen due to their widespread use in industry and regulatory regimes (Dowd and Blake, 2006). Let $\lfloor \cdot \rfloor$ denote the floor function. Then present value of the annuity is given by:

$$P = \frac{1 - 1.05^{-(K-85)}}{0.05}, \tag{31}$$

where $K = \lfloor X \rfloor$ is the curtate age at death random variable.

The theoretical $(\alpha \times 100)\%$ value-at-risk measure for P is

$$\text{VaR}_\alpha(P) = \inf\{p \in \mathbb{R} : F_P(p) > \alpha\}. \tag{32}$$

The $(\alpha \times 100)\%$ conditional tail expectation is defined as

$$\text{TVaR}_\alpha(P) = \mathbb{E}[P|P \geq \text{VaR}_\alpha(P)]. \tag{33}$$

To simulate the curtate future lifetime for a single life, we generate u from the standard uniform distribution and obtain a simulated age at death as

$$K = \sup\left\{ z : \prod_{x=85}^{z} (1 - \hat{q}_x) > U \right\}. \tag{34}$$

We repeat the simulation 1,000,000 times to obtain an empirical distribution of PVs. The simulated expected present value (EPV), $(\alpha \times 100)\%$ VaR, and $(\alpha \times 100)\%$ CTE are then the mean, $(\alpha \times 100)\%$ quantile, and $(\alpha \times 100)\%$ conditional mean of the empirical distribution.

The results are shown in Table 3:

Table 3: Unit annuity: EPV, VaR, and CTE under each method

Confidence Level	**Model**	**EPV($)**	**VaR($)**	**CTE($)**
95%	**Smooth Threshold Life Table**	4.29	9.39	10.62
	Heligman-Pollard	4.31	9.39	10.57
	Coale-Kisker	4.15	8.86	10.02
	Empirical	4.30	9.39	10.64
99.5%	**Smooth Threshold Life Table**	4.29	11.69	12.38
	Heligman-Pollard	4.31	11.27	12.06
	Coale-Kisker	4.15	10.84	11.54
	Empirical	4.30	11.69	12.40
99.9%	**Smooth Threshold Life Table**	4.29	12.46	13.00
	Heligman-Pollard	4.31	12.46	13.07
	Coale-Kisker	4.15	11.69	12.25
	Empirical	4.30	12.46	13.02
99.99%	**Smooth Threshold Life Table**	4.29	13.16	13.56
	Heligman-Pollard	4.31	13.16	13.67
	Coale-Kisker	4.15	12.46	12.86
	Empirical	4.30	13.16	13.49

We also value a theoretical deferred life annuity starting at age 85, that pays $20000 annually to living policyholders in arrears upon survival to age 95. This is a common product in Europe and Asia that allows an individual to hedge their longevity risk (Kartashov, Maurer, Mitchell, and

Rogalla, 2011). We use the same fitted models as for the unit annuity, and compute the same measures. The PV for this annuity is given by

$$PV = \begin{cases} 20000 \times \frac{1-1.05^{-(K-95)}}{0.05 \times 1.05^{10}}, & K \geq 95 \\ 0, & K < 95. \end{cases} \tag{35}$$

The formulae for the VaR and CTE remain as (32) and (33) respectively. The results for this deferred annuity are shown in Table 4.

Table 4: Deferred Annuity: Expected Present Value, Value at Risk, and Conditional Tail Expectation under each method

Confidence Level	Model	EPV($)	VaR($)	CTE($)
95%	**Smooth Threshold Life Table**	4056	33437	57918
	Heligman-Pollard	3942	33437	56894
	Coale-Kisker	3081	22830	45908
	Empirical	3832	33437	58377
99.5%	**Smooth Threshold Life Table**	4056	79357	93113
	Heligman-Pollard	3942	71047	86738
	Coale-Kisker	3081	62321	76399
	Empirical	3832	79357	93502
99.9%	**Smooth Threshold Life Table**	4056	94810	105610
	Heligman-Pollard	3942	94810	106978
	Coale-Kisker	3081	79357	90484
	Empirical	3832	94810	106010
99.99%	**Smooth Threshold Life Table**	4056	108825	116706
	Heligman-Pollard	3942	108825	118947
	Coale-Kisker	3081	94810	102813
	Empirical	3832	108825	115337

The conclusions are similar for both annuity types. The EPVs are similar for the STLT and Heligman-Pollard models, which are close to the empirical EPV. The Coale-Kisker method results in an EPV significantly lower than that of the empirical EPV. The VaRs and CTEs obtained from the STLT are always closer to the empirical VaRs and CTEs than those obtained from the Heligman-Pollard and Coale-Kisker methods. In particular, the Coale-Kisker method is not competitive.

Of note is that the Heligman-Pollard model severely underestimates risk at the 95% and 99.5% confidence levels when compared to the empirical result. This is because of the exponential trend assumed by the model - where empirically mortality rates do not increase so rapidly around the ages of 90-100. The STLT also underestimates the risk at these levels compared to the empirical result, but only slightly.

The last result evident from this analysis is that the Heligman-Pollard model overestimates risk at the very tail of the distribution, evident from the 99.9% and 99.99% VaRs and CTEs. This is again due to the exponential trend of q_x by age, where empirically death rates rise much faster at the very extreme ages (> 105). Again, the STLT also overestimates at the 99.99% level, but not as severely. These results hold for all cohorts, and across both genders (see Figures 22 and 23 in Appendix D).

5 Results for the Dynamic STLT

In this section, we implement time-varying parameters for the smooth threshold life table. Our main steps are as follows:

1. Using death rates from the CBS data for cohorts born in 1893-1901, fit the STLT with time-varying parameters.
2. Using the same data, fit the Cains-Blake-Dowd (CBD) model which is often used to extrapolate death rates into higher ages.
3. Examine the in-sample sum of squared errors (defined by (30)) under each model.
4. Forecast death rates into the future under each model, and compare with realised rates via out-of-sample prediction for the cohorts born in 1902-1908.

The specification of our time-varying model is discussed in the following subsection.

5.1 Exploratory Data Analysis and Model Discussion

To investigate the form our model should take, we start by examining how death rates depend on age for each cohort. The results shown are for females born in the years 1893-1908, for ages 65-99.

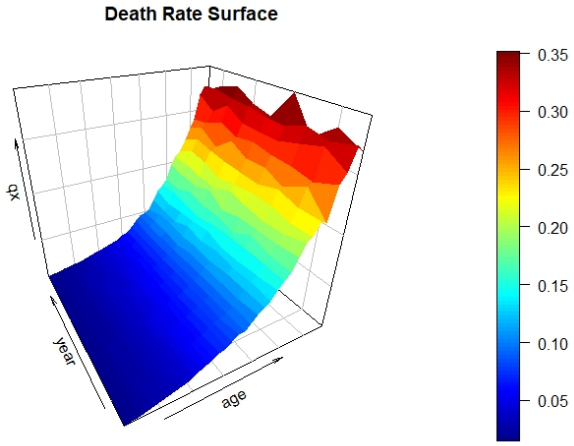


Figure 9: *Female empirical death rate surface*

Death rates tend to increase with age, and slowly decrease with time. This is consistent when examining the fitted STLTs to each cohort, at the lower ages (65-90). At the higher ages (> 90) however, there is no clear pattern. The results shown in Figures 10 - 12 are for females born in the years 1893-1901. Similar findings have also been observed in recent literature; see for example Thatcher (1999).

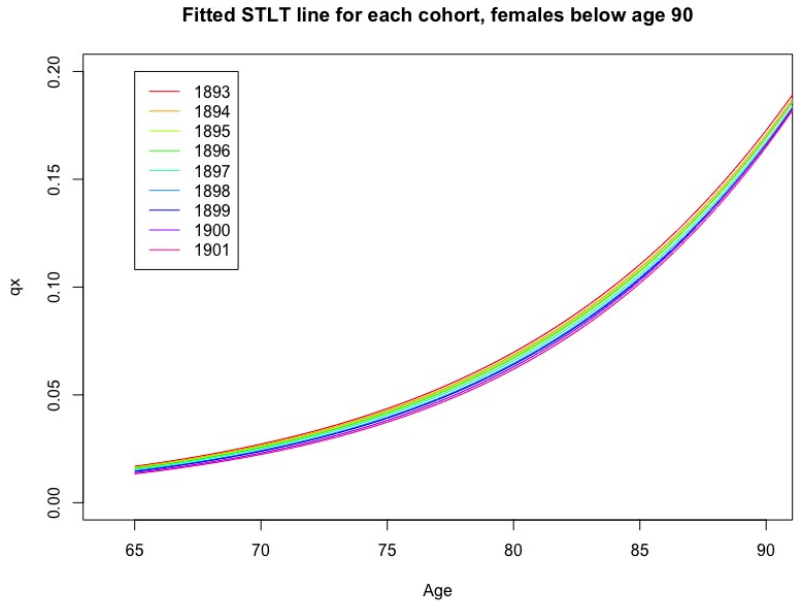


Figure 10: *Fitted STLT lines for each female cohort, shown for ages 90 and below*

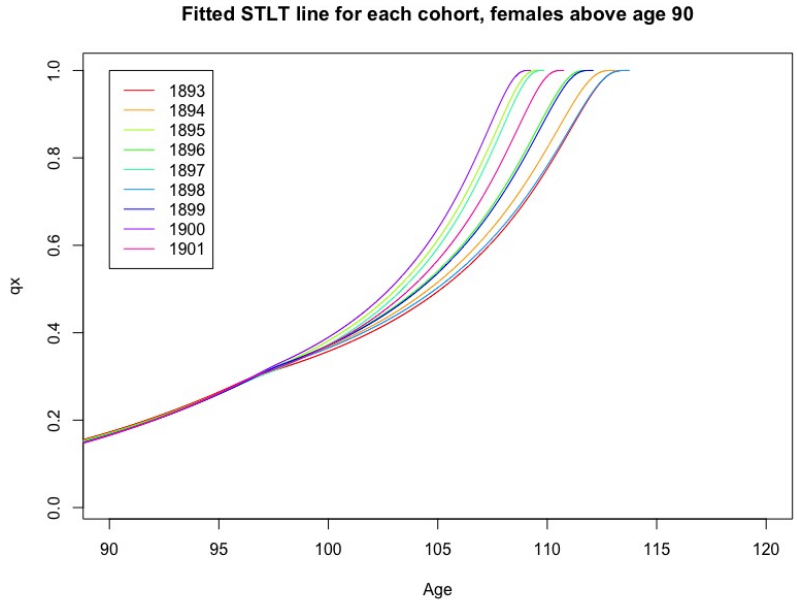


Figure 11: *Fitted STLT lines for each female cohort, shown for ages 90 and above*

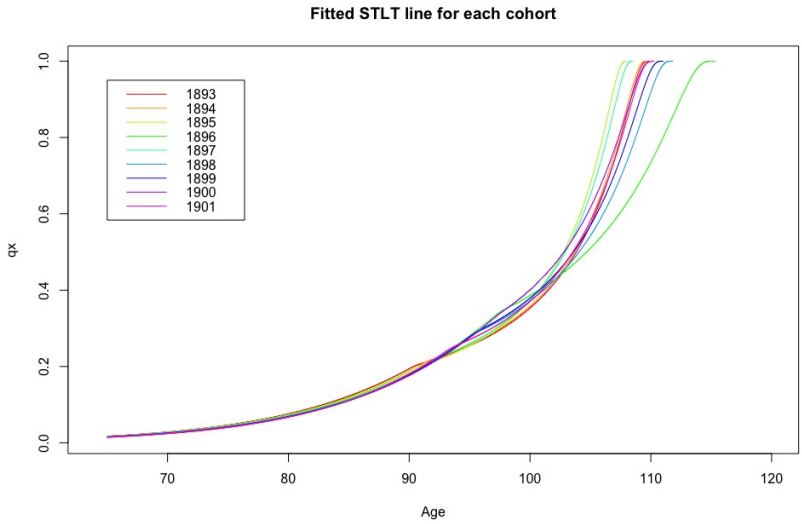


Figure 12: *Fitted STLT lines for each female cohort*

The estimated parameters for each cohort from the STLT model are in Tables 5 - 6. We plot the time-variation of estimated parameters of the 1893-1908 cohorts[4] in Figures 13 and 14 to understand their dynamic patterns. For females, we see that $\hat{B}$ has a noticeable downward trend over cohort, while $\hat{C}$ has an upward trend. For males, however, we observe an increasing trend $\hat{B}$ over the first three cohorts (1893-1895) which levels off over the remaining cohorts, while the reverse is true for $\hat{C}$. That may be due to historical factors such as the impact of the two World Wars. For both genders, we find that $\hat{\gamma}$, $\hat{\theta}$, $\hat{\omega}$ and N do not have discernible trends. This finding is consistent with that of Einmahl, Einmahl, and de Haan (2019).

Table 5: Time-variation of estimated parameters, females

Cohort	$\hat{B}$	$\hat{C}$	$\hat{\theta}$	$\hat{\gamma}$	$\hat{\omega}$	N
1893	0.000031	1.1013	2.82	-0.159	114.74	97
1894	0.000029	1.1019	2.58	-0.160	114.13	98
1895	0.000028	1.1022	2.59	-0.205	110.67	98
1896	0.000026	1.1032	2.58	-0.173	112.95	98
1897	0.000024	1.1039	2.87	-0.207	110.88	97
1898	0.000021	1.1054	2.59	-0.154	114.77	98
1899	0.000020	1.1060	2.59	-0.171	113.10	98
1900	0.000017	1.1081	2.55	-0.208	110.24	98
1901	0.000015	1.1093	2.83	-0.191	111.78	97

[4]Cohorts 1893-1901 are used for model fitting, and cohorts 1902-1908 are used for out-of-sample validation.

Table 6: Time-variation of estimated parameters, males

Cohort	$\hat{B}$	$\hat{C}$	$\hat{\theta}$	$\hat{\gamma}$	$\hat{\omega}$	N
1893	0.000071	1.0945	4.55	-0.221	109.55	89
1894	0.000078	1.0934	3.78	-0.195	110.42	91
1895	0.000086	1.0920	4.20	-0.221	109.00	90
1896	0.000103	1.0895	2.18	-0.176	110.38	98
1897	0.000119	1.0877	2.21	-0.179	110.37	98
1898	0.000104	1.0896	2.15	-0.151	112.27	98
1899	0.000112	1.0885	2.84	-0.179	110.82	95
1900	0.000106	1.0893	2.56	-0.134	115.17	96
1901	0.000118	1.0878	3.10	-0.181	111.12	94

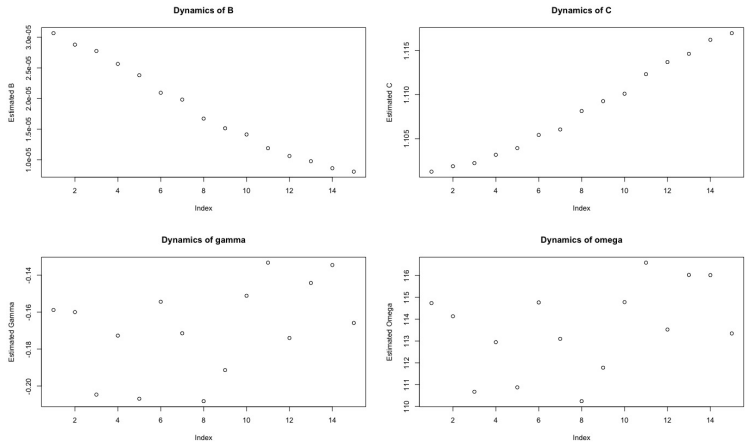


Figure 13: *Time series plots of estimated parameters for the female cohorts. The index corresponds to the cohort, eg. index 1 corresponds to the 1893 cohort.*

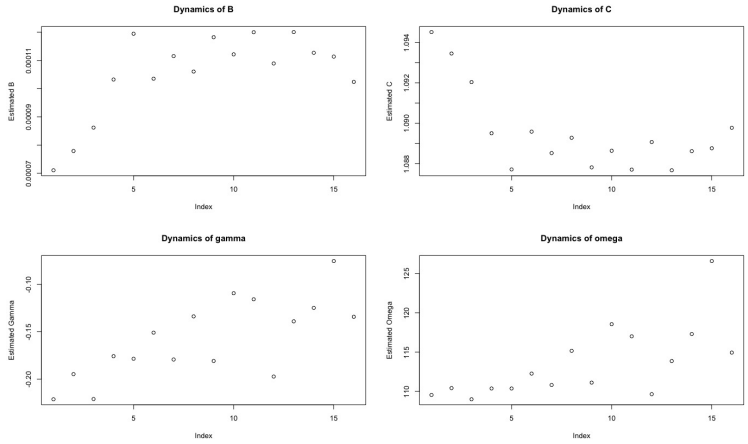


Figure 14: *Time series plots of estimated parameters for the male cohorts. The index corresponds to the cohort, eg. index 1 corresponds to the 1893 cohort.*

Based on these findings, we model B in the STLT as an exponential function of time:

$$B_i = \exp(a + bi), \tag{36}$$

while keeping γ and $\theta = \frac{1}{B_i C_i^N}$ constant through time. Modelling B in this fashion is consistent with the work of Cairns (Cairns, Blake, Dowd, et al., 2011) who model $\log(B)$ as a random walk with drift. In (36), a, b are constants and $i = 1, 2, \ldots 9$ is an index representing the year of birth of a given cohort. Thus $i = 1$ corresponds to the cohort born in 1893, $i = 2$ corresponds to the cohort born in 1894, and so on. The formulation ensures that $B_i > 0$ for all i, and that as B_i decreases, C_i increases in a way such that h_x is always $1/\theta$ at $x = N$.

We call this modified model the dynamic smooth threshold life table (DSTLT). The DSTLT is formulated in the same way as in (12) and (13), with B replaced with $B_i = \exp(a + bi)$ and C replaced with $C_i = (1/(\theta B_i))^{1/N} = \theta^{-N} \exp(-N(a + bi))$; thus, the model is defined by

$$F_i(x) = 1 - \exp\left(\frac{N \exp(a + bi)}{\ln \theta + a + bi} \theta^{-x/N} \exp\left(\frac{-ax - bix}{N}\right) - 1\right) \qquad x \leq N, \tag{37}$$

and

$$F_i(x) = F_\gamma(x) = \begin{cases} 1 - S(N)(1 + \gamma \frac{x-N}{\theta})^{-\frac{1}{\gamma}}, & \gamma > 0,\ x > N \\ 1 - S(N) \exp(-\frac{x-N}{\theta}), & \gamma = 0,\ x > N \\ 1 - S(N)(1 - |\gamma| \frac{x-N}{\theta})^{\frac{1}{|\gamma|}}, & \gamma < 0,\ N < x < N + \frac{\theta}{|\gamma|}. \end{cases} \tag{38}$$

In our model, the force of mortality is continuous at age N through time. Note that under the TLT, modelling time-varying B and C with constant θ and γ would result in the discontinuity at age N being exacerbated through time.

Estimation is performed by finding the a, b, θ and γ that maximize the likelihood,

$$L(a,b,\theta,\gamma;N) = \prod_{i=1}^{9}\left\{\left[\prod_{x=65}^{N-1}\left(\frac{S_i(x)-S_i(x+1)}{S_i(65)}\right)^{d_x}\prod_{x=N}^{\tau-1}\left(\frac{S_i(x)-S_i(x+1)}{S_i(65)}\right)^{d_x}\right]\left(\frac{S_i(\tau)}{S_i(65)}\right)^{l_\tau}\right\}, \tag{39}$$

for a given N, with $\hat{N}$ subsequently chosen as the value of N that results in the highest maximized likelihood. This N is assumed to be constant across all cohorts.

The DSTLT model indicates that the predicted q_x for $x \leq N$ will decrease over time, while q_x for $x > N$ will remain constant over time. This is consistent with our empirical evidence and the existing literature; for example see Thatcher (1999). Although the probabilities of dying at lower ages in the 19th century were much higher than they are today, there has been little change in the probabilities of dying at the higher ages.

Fitting this model, we obtain the parameter estimates in Tables 7 and 8, with the estimated threshold age $N = 98$ for females and $N = 96$ for males.

Table 7: Estimated parameters for the DSTLT, females

Parameter	Estimate	Standard Error	95% Confidence Interval
a	-10.26	0.023	(-10.31, -10.22)
b	-0.085	0.0033	(-0.092, -0.079)
θ	2.58	0.0087	(2.56, 2.60)
γ	-0.174	0.0057	(-0.185, -0.163)

Table 8: Estimated parameters for the DSTLT, males

Parameter	Estimate	Standard Error	95% Confidence Interval
a	-9.29	0.024	(-9.34, -9.24)
b	0.023	0.0033	(0.017, 0.029)
θ	2.57	0.010	(2.55, 2.59)
γ	-0.156	0.0074	(-0.170, -0.141)

To determine whether our dynamic modification is a significant improvement over the static model, we test the hypothesis

$$H_0 : b = 0 \quad \text{vs} \quad H_1 : b \neq 0,$$

using a likelihood ratio test. The test statistic (TS) is given by

$$\text{TS} = -2 \ln \left(\frac{L_D(\hat{a}, \hat{b}, \hat{\theta}, \hat{\gamma}; N)}{L_S(\hat{a}, \hat{\theta}, \hat{\gamma}; N)} \right),$$

where $L_D(\hat{a}, \hat{b}, \hat{\theta}, \hat{\gamma}; N)$ is the likelihood as defined in (39) evaluated at the MLEs of the parameters, and $L_S(\hat{a}, \hat{\theta}, \hat{\gamma}; N)$ denotes the likelihood of the STLT evaluated at its MLEs using data from all years in the sample. This test statistic is asymptotically χ^2_1 distributed under the null hypothesis. The observed test statistic is 657.76, far exceeding the 0.001 quantile of 10.83 for a χ^2_1. We thus reject H_0 and conclude that there is a significant time-varying pattern. The same conclusion is obtained when using males data; for males however, the trend for the parameter B is weakly increasing with cohort.

5.2 Comparison with CBD model

The CBD model is fitted from age 65 to 89, and extrapolated into the higher ages. This is the suggested method in Currie (2011) for modelling advanced age mortality with the CBD model. The predicted q_x for the cohorts born in 1902-1908 under the DSTLT and CBD models are shown in Figures 15 and 16.

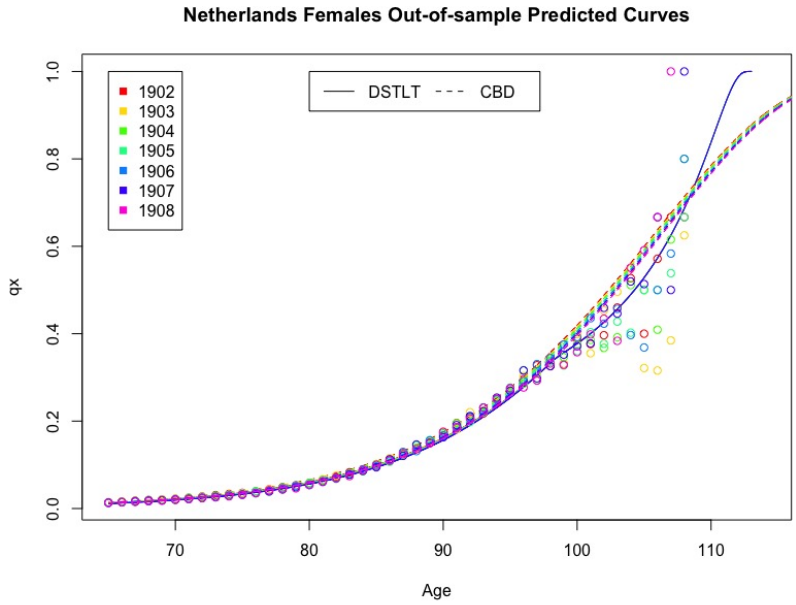


Figure 15: *Out-of-sample predicted DSTLT and CBD curves for the 1902-1908 female cohorts*

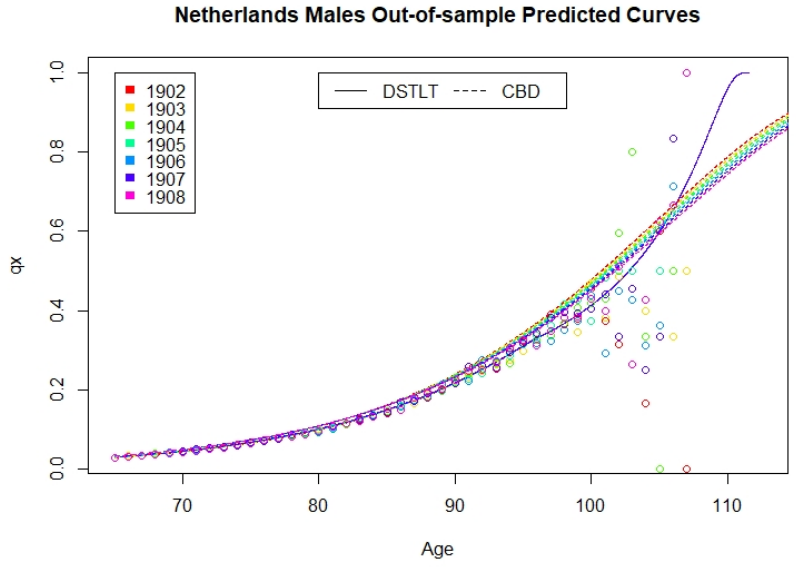


Figure 16: *Out-of-sample predicted DSTLT and CBD curves for the 1902-1908 male cohorts*

For females, the DSTLT predictions are very close to the CBD predictions for ages below the estimated threshold age 98. From ages 98 to 106, the DSTLT predicts a lower probability of death

than the CBD model does, consistent with the data. After age 106 the DSTLT predicts a higher probability of death than the CBD model, again consistent with the data - though based on very scant data for these cohorts. For males, the interpretation is similar, but with the threshold age estimated to be 96 years. To make a formal comparison, we summarize the in- and out-of-sample prediction SSEs in Tables 9 - 12.

Table 9: In-sample fitting errors under the DSTLT and CBD models, females

	Below age 98		Above age 98		All ages	
Cohort	DSTLT	CBD	DSTLT	CBD	DSTLT	CBD
1893	0.001	0.028	0.262	0.660	0.262	0.688
1894	0.000	0.020	0.125	0.308	0.126	0.328
1895	0.000	0.017	0.628	0.760	0.628	0.777
1896	0.000	0.012	0.093	0.194	0.094	0.206
1897	0.000	0.012	0.578	0.663	0.578	0.674
1898	0.000	0.013	0.170	0.353	0.171	0.366
1899	0.001	0.008	0.287	0.411	0.288	0.420
1900	0.001	0.006	0.604	0.643	0.605	0.649
1901	0.001	0.010	0.121	0.203	0.121	0.213

Table 10: Out-of-sample prediction errors under the DSTLT and CBD models, females

	Below age 98		Above age 98		All ages	
Cohort	DSTLT	CBD	DSTLT	CBD	DSTLT	CBD
1902	0.001	0.009	0.030	0.178	0.031	0.187
1903	0.001	0.006	0.226	0.445	0.227	0.452
1904	0.001	0.005	0.053	0.214	0.054	0.220
1905	0.001	0.006	0.046	0.210	0.047	0.215
1906	0.001	0.004	0.064	0.200	0.066	0.204
1907	0.002	0.003	0.105	0.171	0.106	0.174
1908	0.002	0.002	0.132	0.148	0.134	0.151

Table 11: In-sample fitting errors under the DSTLT and CBD models, males

	Below age 96		Above age 96		All ages	
Cohort	DSTLT	CBD	DSTLT	CBD	DSTLT	CBD
1893	0.001	0.043	0.262	0.609	0.262	0.652
1894	0.000	0.032	0.125	0.713	0.126	0.745
1895	0.000	0.050	0.628	0.428	0.628	0.478
1896	0.000	0.043	0.093	0.783	0.094	0.825
1897	0.000	0.032	0.578	0.509	0.578	0.541
1898	0.000	0.029	0.170	0.439	0.171	0.468
1899	0.001	0.027	0.287	0.220	0.288	0.247
1900	0.001	0.025	0.604	0.926	0.605	0.950
1901	0.001	0.038	0.121	0.839	0.121	0.877

Table 12: Out-of-sample prediction errors under the DSTLT and CBD models, males

	Below age 96		Above age 96		All ages	
Cohort	DSTLT	CBD	DSTLT	CBD	DSTLT	CBD
1902	0.001	0.040	0.726	1.227	0.726	1.267
1903	0.000	0.044	0.197	0.573	0.197	0.616
1904	0.001	0.043	0.608	0.900	0.609	0.943
1905	0.000	0.047	0.186	0.431	0.187	0.478
1906	0.001	0.041	0.225	0.596	0.226	0.637
1907	0.001	0.043	0.295	0.625	0.296	0.668
1908	0.001	0.043	0.158	0.424	0.159	0.467

The DSTLT model outperforms the CBD model across all cohorts in the empirical dataset in terms of out-of-sample forecasting for both males and females. For in-sample fitting, the DSTLT performs worse than the CBD model only for the male cohorts of 1895, 1897, and 1899. The discrepancy in these cases invariably occurs at extreme ages. But at extremely high ages, the number of survivors is in single digits and observed probabilities of death may even be 0. For example, the observed q_{107} for the 1899 male cohort is 0. The SSE measure then overwhelmingly favours whichever model predicts lower probabilities of death for that age.

For both in-sample fitting and out-of-sample prediction errors (see Figures 24 - 27 in Appendix E), the CBD model and DSTLT perform similarly for ages under 90. From ages 90-105, the CBD model overestimates probabilities of death for almost all cohorts. For the very high ages past age 105, the CBD model then underestimates probabilities of death. The DSTLT also overestimates and underestimates probabilities of death similarly, but not as much as the CBD model.

6 Discussion

6.1 Implications of our model

In the literature, there have been three distinct philosophies regarding the tail behaviour of the lifetime distribution. The first proposes that the conditional probability of dying within one year reaches 1 at a finite age and thus there is a fixed upper limit to the length of human life (e.g., Vincent 1951; Gbari, Poulain, Dal, and Denuit 2017; Ferreira and Huang 2018; Einmahl et al. 2019). The second argues that this probability remains below 1 at finite ages, but nevertheless asymptotically tends to 1 (Heligman and Pollard 1980; Gompertz 1825). This is the "life is unlimited but short" scenario (Rootzén and Zholud, 2017). The final view suggests that the conditional probability of dying within one year may asymptote to a limit less than 1 (e.g., Thatcher 1999). The TLT and STLT models provide an objective way of responding to this question. If $\gamma < 0$, then $q_x = 1$ at $x = N - \theta/\gamma$. When $\gamma = 0$, the tail is exponential, so that q_x equals some constant less than one for all ages above age N. Finally, $q_x \to 0$ as $x \to \infty$ is also allowed, corresponding to the case where $\gamma > 0$. Although the second philosophy is not explicitly captured, it is implicit when γ is both negative and very close to zero.

The majority of our analyzed cohorts belong to the first case. This is consistent with the notion that humans do not live forever, possibly constrained by factors such as telomere attrition (Steenstrup et al., 2017) which places a finite limit on the number of times a cell can undergo mitosis.

However, many researchers have also observed a late-life mortality deceleration effect - that is, the rate of mortality increase being slower than exponential at advanced ages - and in some cases plateauing at a constant rate (e.g. Greenwood and Irwin, 1939; Gavrilov and Gavrilova, 2011 ; Barbi, Lagona, Marsili, Vaupel, and Wachter, 2018). This effect is usually understood to imply a model with an infinite human life span, which seemingly contradicts the fact that humans do not live forever. The DSTLT model reconciles these two observations by allowing an advanced age

late-life mortality acceleration *after* a late-life mortality deceleration effect, the former of which leads to a finite limit to the human life span. The effect is inherent in the structure of the STLT and DSTLT models. To see this, we can examine the log force of mortality under the Gompertz distribution and GPD.

Let ${}_1h_x$ and ${}_2h_x$ denote the hazard function at age x for the Gompertz distribution and GPD respectively (see Appendix B for derivations of these). Then

$$\begin{aligned} &\ln {}_1h_x = x \ln C + \ln B \\ &\ln {}_2h_x = -\ln(\theta + \gamma(x - N)) \\ \Rightarrow &\frac{d \ln {}_1h_x}{dx} = \ln C \\ &\frac{d \ln {}_2h_x}{dx} = \frac{-\gamma}{(\theta + \gamma(x - N))}. \end{aligned}$$

Setting the two derivatives to be equal, we solve for x:

$$\begin{aligned} \ln C =& \frac{-\gamma}{(\theta + \gamma(x - N))} \\ \frac{-\ln C}{\gamma} =& \frac{1}{\theta + \gamma(x - N)} \\ \frac{-\gamma}{\ln C} =& \theta + \gamma(x - N) \\ x^* =& \frac{\frac{-\gamma}{\ln C} - \theta}{\gamma} + N \\ =& N - \frac{1}{\ln C} - \frac{\theta}{\gamma}. \end{aligned} \tag{40}$$

So before this age x^*, the force of mortality under the GPD increases slower than that of the Gompertz. The reverse is true after age x^*. When $\gamma < 0$, as is always the case in our CBS data, deceleration will be observed after the threshold age N if and only if $\theta \ln C > |\gamma|$. This is indeed always the case in our data.

The other key feature of the DSTLT is that the function C_i, the age effect, increases as B_i, or baseline mortality, decreases - that is, C_i is a strictly decreasing function of B_i - such that the force of mortality of different cohorts converges at the threshold age. In the DSTLT model, the force of

mortality of different cohorts is the same at and above the threshold age - that is, $h_i(x) = h(x)$ for all $i = 1, 2, \ldots, 9$ and $x \geq N$; in particular, $h_i(x) = 1/\theta$ at $x = N$ for all i. This explicitly captures the *compensation law of mortality* (Gavrilov and Gavrilova, 1979). The law states that for a given species, differences in death rates between different sub-populations decrease with age - because higher initial death rates are accompanied by a lower rate of mortality increase with age - such that mortality rates for all sub-populations (or cohorts) converge at some high age. The age at which this is presumed to happen is termed the "species-specific life span", which is captured by the threshold age N in the DSTLT model. The human species-specific life span (i.e. the threshold age N) is estimated as 98 years for females and 96 years for males using our data, which are close to the estimate of around 95 years by Gavrilov and Gavrilova (1991). Consider also the mortality trends of both genders (see Figure 28 in Appendix F). Although females in our data have consistently lower mortality rates at the non-extreme ages than males, there is no significant difference between them at the extreme ages. This is also consistent with the compensation law of mortality.

The significance of the smoothing constraint should also be emphasized here - it is pivotal in generalizing the TLT model to the DSTLT model. If we simply modelled parameters B and C through time in the TLT model, the force of mortality would not be continuous at age N. Furthermore, the size of the discontinuity will eventually get worse with time. The fact that the DSTLT model satisfies the compensation law of mortality also relies on the smoothing constraint, in that it forces C_i to be the correct reciprocal function of B_i.

Owing to its form, the DSTLT model actually satisfies all three laws of biodemography[5]: *Gompertz-Makeham law of mortality, compensation law of mortality* and *late-life mortality deceleration*, while simultaneously allowing a finite limit to human life span. To the best of our knowledge, no other model in the existing literature has all these properties.

[5] https://en.wikipedia.org/wiki/Biodemography_of_human_longevity

6.2 Precision of estimation

The MLE approach uses the information matrix to assess precision of estimation of parameters. Smith (1985) shows that the information matrix for the GPD is finite if and only if $\gamma > -1/2$. This is a necessary condition for our estimators to be efficient, consistent, and asymptotically satisfy the Cramer-Rao lower bound. This condition is satisfied for all the cohorts we fit our models to.

Alternatively, estimation can be performed using other techniques such as the method of moments or the method of probability weighted moments. However, these two methods are not feasible when $\gamma < 0$ (Dupuis and Field, 1998). These two methods are therefore not relevant to our application as the lifetime distribution (GPD in our model) only has an upper bound when $\gamma < 0$. Peng and Welsh (2001) also outlined a 'method of medians', which is robust. The estimates are obtained by equating the sample medians of the components in the score function to the population median for real data. This method is technically challenging given our piecewise model.

Another issue to consider is that exposure-to-risk is a decreasing function with respect to age, so the d_x and l_x used in fitting the Gompertz part of the distribution will typically be much higher than that of the Generalised Pareto distribution. A consequence of this is that the total likelihood plateaus for high values of N (see Figure 17), because even if the fit is poor at the higher ages, the smaller numbers of exposure-to-risk mean that the total likelihood will not change much.

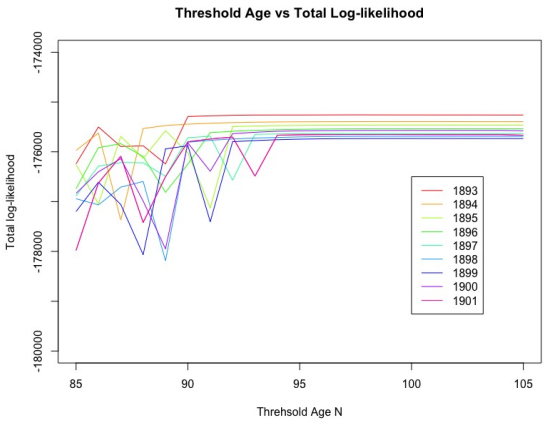


Figure 17: *Total log-likelihood of the STLT fit as a function of N, females*

7 Conclusion

The threshold life table is a method that utilises extreme value theory by combining the Gompertz distribution with the generalised Pareto distribution to model advanced-age mortality. In this thesis, we rectify the shortcomings of the threshold life table by proposing two modifications. One is to impose a smoothing constraint (the STLT model). The other is to introduce a time-varying component (the DSTLT model). We analyze the two modifications using a high quality data set constructed by combining data from the Human Mortality Database and Statistics Netherlands.

7.1 Summary of Findings

The findings of this thesis may be summarised as follows.

1. The smoothing constraint removes the discontinuity in the force of mortality that occurs at threshold age N in the threshold life table. Furthermore, it provides a functional link between the Gompertz and GPD distributions, which tends to make the extreme age modelling of the STLT more robust than that of the TLT model.

2. The inclusion of high quality data (HMD+CBS) significantly improves the estimation precision of the important parameter γ, which bears on the question of the existence of an upper limit to the human life span.

3. Estimated ω is greatly data dependent. Using the HMD data to fit the smooth threshold life table results in a reduced $\hat{\omega}$ across all cohorts analyzed than if the model is fitted to the combined data.

4. The STLT fits the empirical data better than the existing TLT, Helligman-Pollard and Coale-Kisker methods, and quantifies longevity risk more accurately.

5. The dynamic smooth threshold life table predicts future death rates more accurately than

the Cairns-Blake-Dowd model, especially for the extreme ages.

6. The DSTLT simultaneously allows a finite limit to the human lifespan and captures all three general biodemographic laws of survival. The model therefore opens up a new perspective on the discussion of the limits of human longevity. Namely, our findings assert both late-life mortality deceleration and a finite limit to human lifespan.

7.2 Limitations

One major limitation in this study is the availability of data. The data provided by the Statistics Netherlands only contains age and year at death, and not the number alive at each age and year. Thus the number alive needs to be calculated by assuming all individuals for a given cohort has died by the end of the observation period (the extinct cohorts method). We are only able to reasonably do this for up to 15 cohorts, resulting in a large portion of the data set being unused. This is especially problematic when analyzing the DSTLT as there are only a small number of observations available to examine the time-variation in parameters.

Another problem not fixed by our proposed modifications is that the standard error of N is still somewhat large. The same problem also persists for $\hat{\gamma}$, though lessened in severity by combining the high quality CBS data.

One final point of note is that using a generalised Pareto distribution to model the conditional exceedances assumes that the distribution of the scaled maximum converges to a non-degenerate distribution.

7.3 Further Research

One avenue of potential future research is in an improved method of selecting N. The use of cross-validation is one approach that could be taken here, aiming to minimize the in-sample sum

of squared errors or some other objective function. Another approach could be to use a different estimation scheme, such as maximum weighted likelihood. The weighted likelihood method can also be applied in the estimation of γ, in attempt to increase precision of the estimator by trading increased bias for a decreased variance.

The DSTLT can also be further extended to model not just advanced-age mortality for a single population, but for multiple populations. For example, we could enforce a constant θ and γ assumption through the time and population dimension to maintain agreement with the compensation law of mortality, and let the parameters a and b vary by, for example, country or gender. A further natural extension to this concept would be to model a and b with covariates.

Finally, the models proposed in this thesis is only suitable for the mortality modelling of adult ages. In future research, we would like to investigate a new model that can generate life tables for the entire life span.

Appendices

A Data Validation

Table 13: Differences in the number of age and year-specific deaths between the HMD and CBS datasets for Males (left) and Females (right)

Year/Age	93	94	95	96	97	98	99
1986	3	0	4	2	2	1	1
1987	2	3	0	0	1	1	0
1988	1	2	2	2	1	1	0
1989	3	2	1	0	1	0	0
1990	1	3	1	3	3	1	0
1991	2	1	0	1	0	1	0
1992	5	3	0	1	0	1	0
1993	1	1	0	1	2	2	0
1994	1	2	2	1	1	1	0
1995	3	1	1	2	1	0	4
1996	0	2	1	0	1	0	1
1997	2	2	0	4	2	0	1
1998	2	3	1	1	1	0	2
1999	3	3	1	2	2	0	0
2000	0	0	1	1	0	0	1
2001	3	1	2	0	1	0	0
2002	2	2	1	2	3	0	0
2003	3	0	2	0	0	0	0
2004	0	3	3	0	0	1	0
2005	6	1	0	0	0	0	0
2006	1	0	0	0	0	1	0
2007	0	2	0	1	1	0	0
2008	0	0	0	1	0	1	0
2009	0	0	0	1	0	1	0
2010	1	0	1	0	0	1	0
2011	3	0	1	0	1	1	0
2012	1	0	1	0	2	2	0
2013	5	0	0	0	0	0	0
2014	1	2	0	0	0	0	0
2015	1	4	1	0	0	0	1

Year/Age	93	94	95	96	97	98	99
1986	4	2	1	2	0	1	0
1987	2	1	2	3	0	3	0
1988	1	0	1	0	1	1	1
1989	3	5	4	1	3	0	1
1990	11	1	1	0	0	2	0
1991	2	1	5	5	5	0	1
1992	2	1	2	1	0	1	0
1993	2	0	1	1	1	2	2
1994	1	0	3	2	2	3	2
1995	8	2	6	4	2	1	17
1996	1	4	1	1	2	3	4
1997	1	8	7	1	6	1	1
1998	0	5	6	4	4	1	3
1999	4	1	1	0	2	1	1
2000	0	6	4	2	3	1	2
2001	1	0	6	4	1	0	0
2002	2	1	4	0	1	1	2
2003	3	2	3	1	1	1	1
2004	2	1	2	2	1	2	1
2005	0	1	1	0	0	2	1
2006	0	7	2	5	5	0	0
2007	0	1	7	0	1	0	1
2008	1	0	3	1	1	0	1
2009	8	1	1	0	1	0	1
2010	3	4	5	2	1	2	1
2011	4	3	8	1	1	2	1
2012	3	3	2	1	1	2	1
2013	8	0	1	1	3	0	0
2014	3	9	1	1	2	2	0
2015	2	1	11	4	5	2	2

Table 14: Number of deaths in the HMD database in each year × age-group cell for Males (left) and Females (right)

Year/Age	93	94	95	96	97	98	99
1986	438	392	276	175	142	107	66
1987	446	378	273	190	129	72	72
1988	427	359	284	221	154	96	62
1989	466	349	283	227	158	112	64
1990	458	368	290	187	144	116	64
1991	477	399	309	207	172	126	58
1992	487	378	280	210	148	107	78
1993	514	409	273	215	153	117	71
1994	503	365	276	242	144	118	76
1995	473	397	281	217	158	121	64
1996	517	424	255	202	147	81	67
1997	491	396	314	226	147	105	59
1998	528	358	308	225	152	93	70
1999	536	409	310	253	167	112	66
2000	478	422	321	218	166	96	82
2001	530	411	327	236	153	129	63
2002	531	437	309	215	168	110	63
2003	596	439	323	228	150	123	77
2004	543	420	308	208	168	110	74
2005	620	450	336	241	159	107	76
2006	642	475	331	230	148	114	64
2007	628	480	339	246	163	122	71
2008	663	528	373	272	162	133	79
2009	637	496	391	270	163	117	84
2010	654	533	384	313	178	147	86
2011	677	474	389	291	217	138	91
2012	702	607	446	290	231	158	115
2013	920	550	452	307	239	158	108
2014	955	719	414	330	231	151	111
2015	1056	786	620	339	237	182	105

Year/Age	93	94	95	96	97	98	99
1986	907	683	557	392	295	204	140
1987	978	706	562	454	330	214	136
1988	986	825	592	495	356	238	165
1989	1140	899	675	544	398	281	185
1990	1159	953	751	533	415	269	177
1991	1148	964	750	584	431	294	194
1992	1219	1013	752	595	456	329	201
1993	1404	1120	912	683	505	344	239
1994	1356	1120	881	680	482	356	223
1995	1397	1159	872	693	518	362	218
1996	1459	1219	924	716	530	381	253
1997	1536	1278	1017	720	548	374	262
1998	1557	1262	1029	824	568	409	265
1999	1669	1368	1066	830	643	462	280
2000	1645	1386	1104	796	589	430	276
2001	1746	1457	1146	965	585	455	294
2002	1793	1463	1156	878	654	474	287
2003	1800	1488	1201	989	706	479	323
2004	1714	1390	1185	878	638	492	311
2005	1737	1417	1157	949	671	490	352
2006	1864	1545	1233	951	705	538	357
2007	1794	1519	1238	894	724	499	356
2008	1912	1666	1333	1024	732	606	413
2009	1797	1620	1321	1044	749	572	397
2010	1906	1567	1346	1121	800	582	396
2011	1880	1590	1337	1082	840	604	441
2012	1846	1681	1454	1139	942	668	476
2013	2269	1631	1477	1158	912	753	510
2014	2361	1985	1385	1110	939	717	466
2015	2522	2197	1713	1218	979	706	491

B Derivation of smoothness constraint

Given the threshold life table defined by

$$F(x) = 1 - \exp\left(-\frac{B}{\ln C}(C^x - 1)\right), \qquad x \leq N \tag{41}$$

and

$$F(x) = 1 - S(N)\left(1 + \gamma\left(\frac{x-N}{\theta}\right)\right)^{-1/\gamma}, \qquad x > N, \tag{42}$$

we want $h_1(N) = h_2(N)$ where h_1 is the hazard function corresponding to (41) and h_2 the hazard function corresponding to (42).

For h_1, we have

$$\begin{aligned} f(x) = F'(x) &= -\exp\left(-\frac{B}{\ln C}(C^x - 1)\right) \cdot \frac{-B}{\ln C} \cdot C^x \ln C \\ &= BC^x \exp\left(-\frac{B}{\ln C}(C^x - 1)\right), \end{aligned}$$

so that

$$\begin{aligned} h_1(x) &= \frac{f(x)}{1-F(x)} \\ &= \frac{BC^x \exp\left(-\frac{B}{\ln C}(C^x-1)\right)}{\exp\left(-\frac{B}{\ln C}(C^x-1)\right)} \\ &= BC^x. \end{aligned}$$

Similarly, for h_2 we have

$$\begin{aligned} f(x) = F'(x) &= \frac{\rho}{\gamma}\left(1+\gamma\left(\frac{x-N}{\theta}\right)\right)^{-\left(\frac{1+\gamma}{\gamma}\right)} \cdot \frac{\gamma}{\theta} \\ &= \frac{\rho}{\theta}\left(1+\gamma\left(\frac{x-N}{\theta}\right)\right)^{-\left(\frac{1+\gamma}{\gamma}\right)}, \end{aligned}$$

so that

$$\begin{aligned} h_2(x) &= \frac{f(x)}{1-F(x)} \\ &= \frac{\frac{\rho}{\theta}\left(1+\gamma\left(\frac{x-N}{\theta}\right)\right)^{-\left(\frac{1+\gamma}{\gamma}\right)}}{\rho\left(1+\gamma\left(\frac{x-N}{\theta}\right)\right)^{-1/\gamma}} \\ &= \frac{1}{\theta}\left(1+\gamma\left(\frac{x-N}{\theta}\right)\right)^{-1} \\ &= (\theta+\gamma(x-N))^{-1} \\ &= \frac{1}{\theta+\gamma(x-N)}, \quad \gamma \in \mathbb{R}. \end{aligned}$$

setting $h_1(N) = h_2(N)$, we have

$$\begin{aligned} &\frac{1}{\theta+\gamma(x-N)} = BC^x \qquad \text{at } x = N \\ &\Rightarrow \theta = \frac{1}{BC^N}. \end{aligned}$$

So the distribution to be fitted is as given in (21) and (22).

C Fitted TLT and STLT for All Cohorts

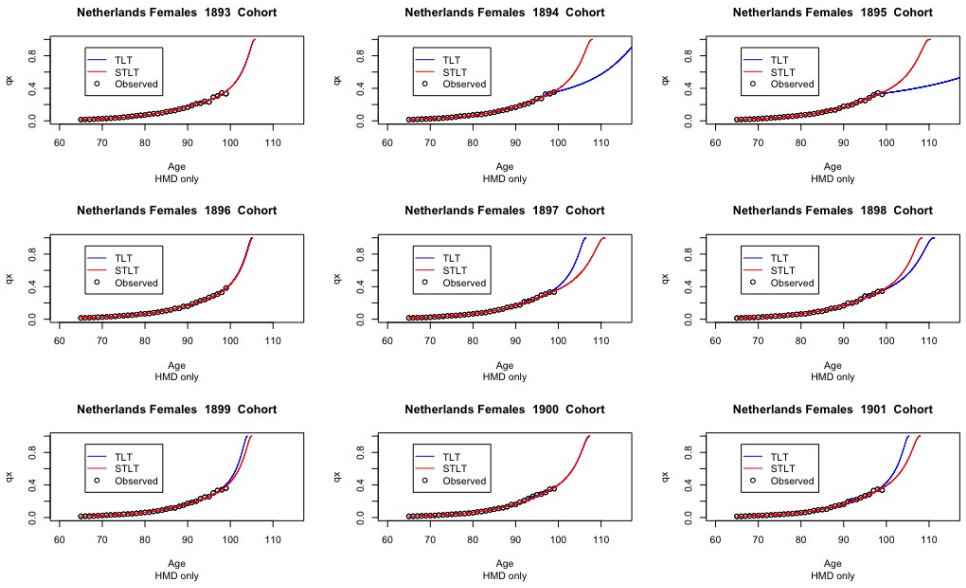


Figure 18: *Fitted TLT and STLT for all cohorts; HMD dataset only, females*

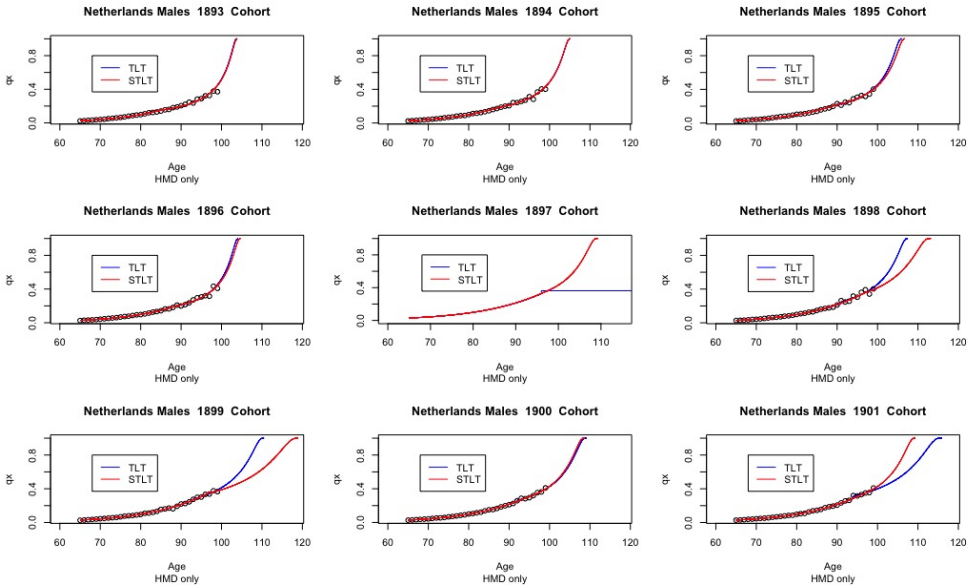


Figure 19: *Fitted TLT and STLT for all cohorts; HMD only, males*

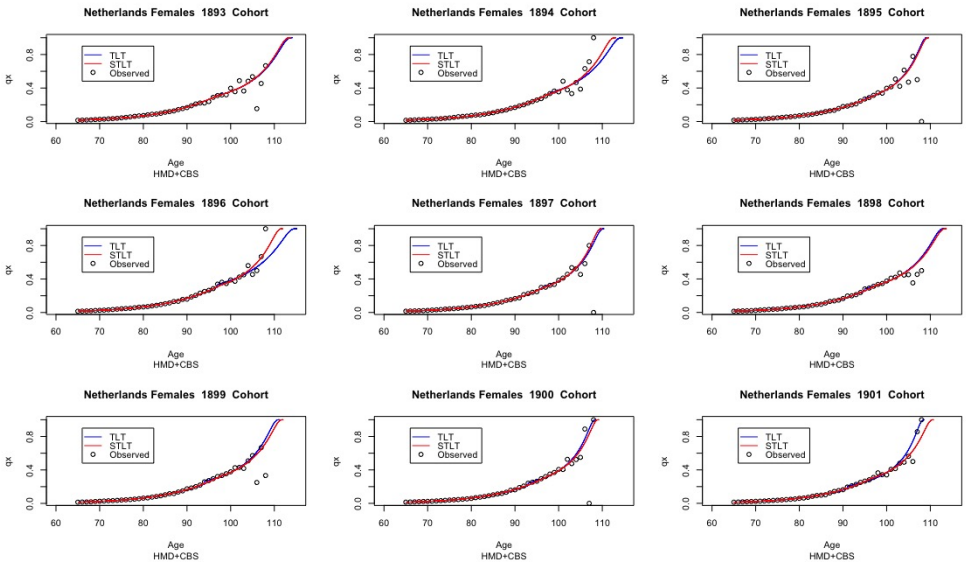


Figure 20: *Fitted TLT and STLT for all cohorts; HMD+CBS dataset, females*

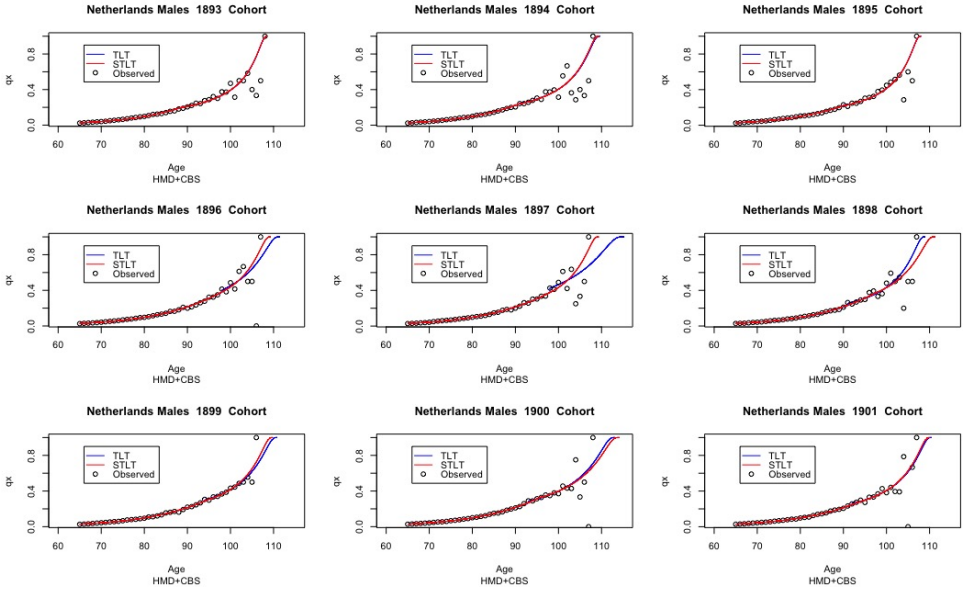


Figure 21: *Fitted TLT and STLT for all cohorts; HMD+CBS dataset, males*

D Heligman-Pollard, Coale-Kisker and STLT Comparisons for All Male and Female Cohorts

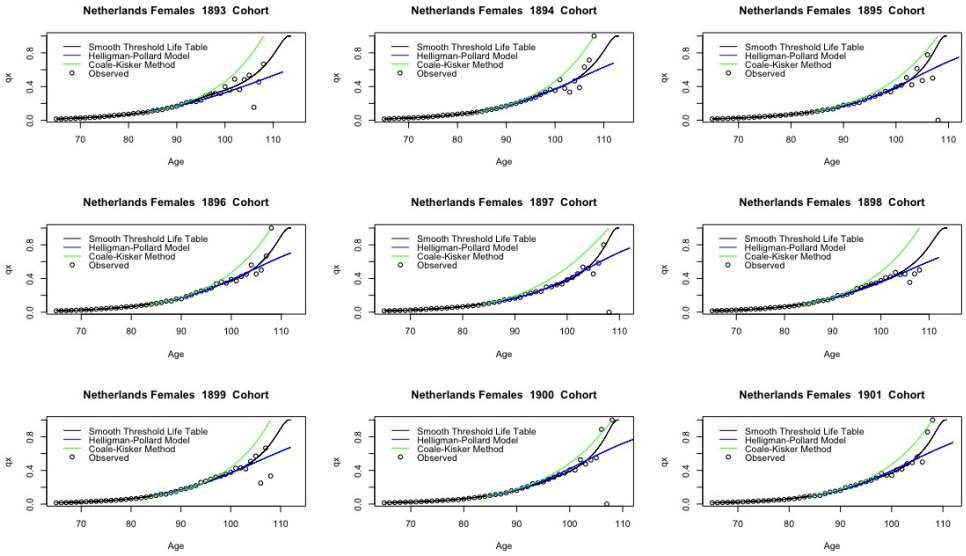


Figure 22: *Fitted STLT, Heligman-Pollard and Coale-Kisker models for all cohorts, females*

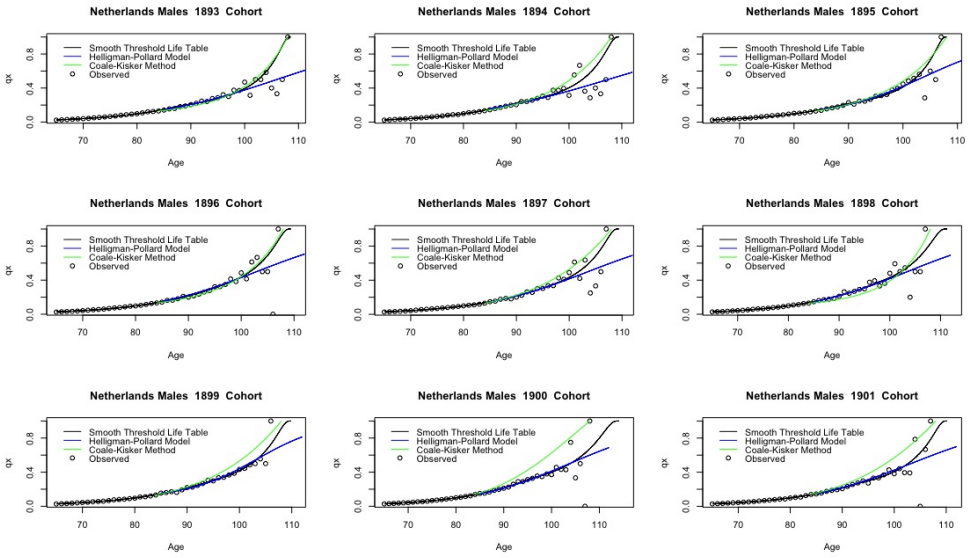


Figure 23: *Fitted STLT, Heligman-Pollard and Coale-Kisker models for all cohorts, males*

E Fitting and Prediction Errors of DSTLT and CBD Models

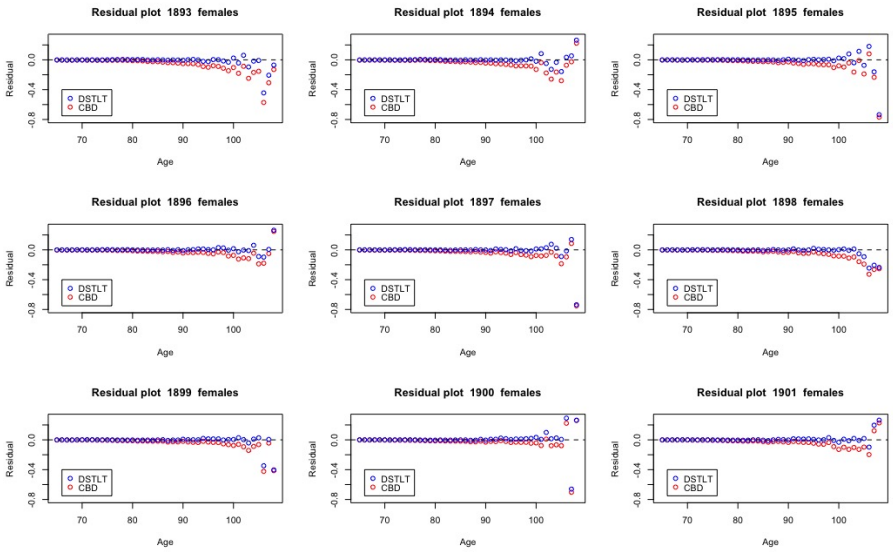


Figure 24: *DSTLT and CBD model residuals, females. Blue points correspond to the DSTLT; red points correspond to the CBD model*

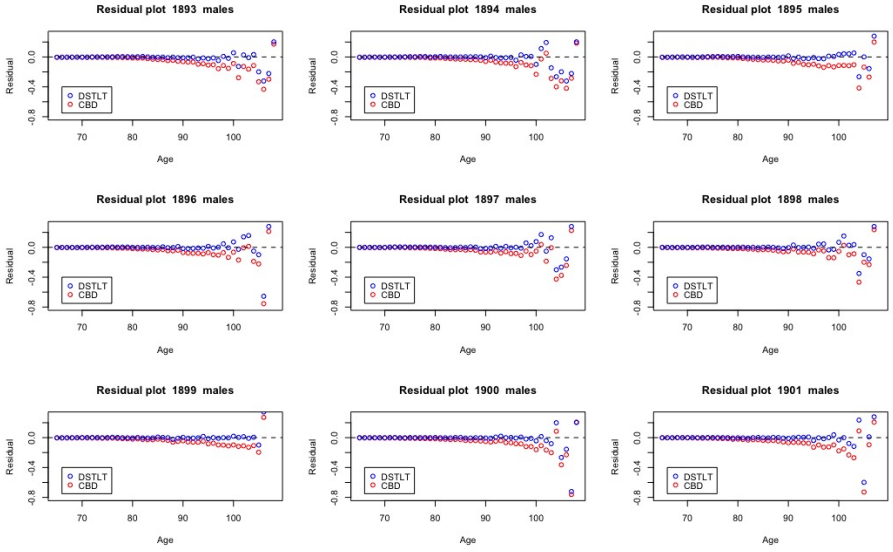


Figure 25: *DSTLT and CBD model residuals, males. Blue points correspond to the STLT; red points correspond to the CBD model*

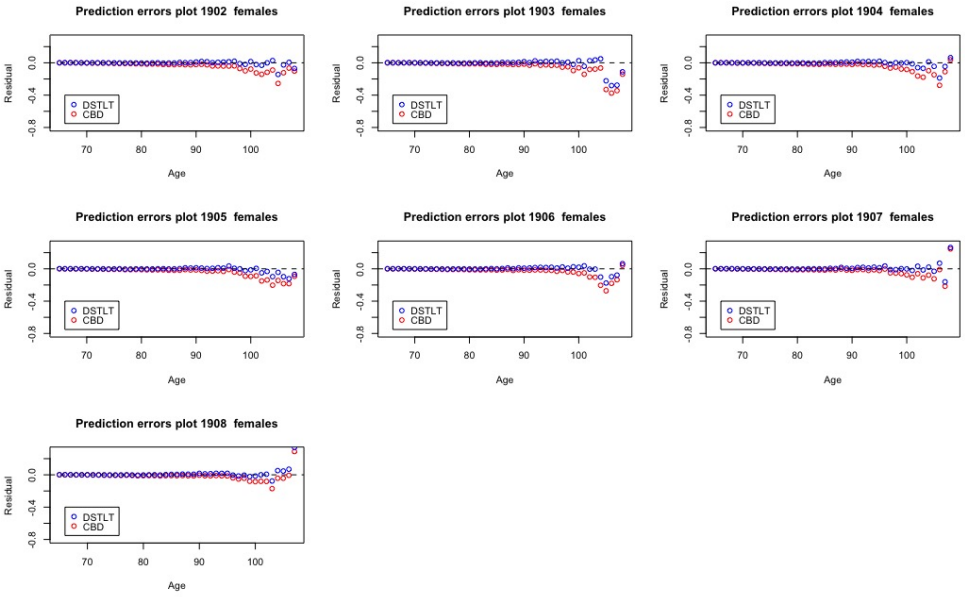


Figure 26: *Prediction errors under the DSTLT and CBD model, females. Blue points correspond to the DSTLT; red points correspond to the CBD model*

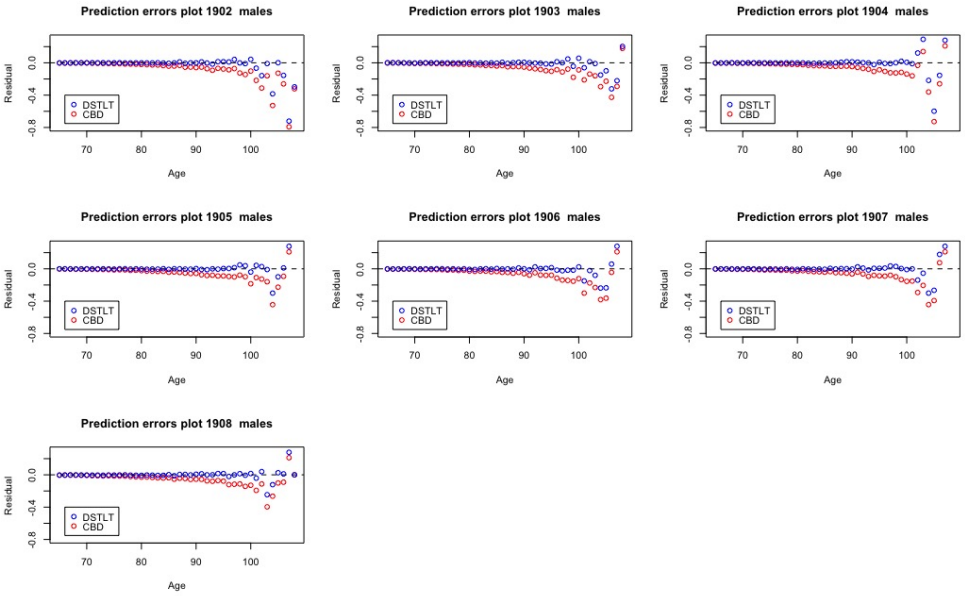


Figure 27: *Prediction errors under the DSTLT and CBD model, males. Blue points correspond to the STLT; red points correspond to the CBD model*

F STLT fit male vs female comparison, 1893-1901 cohorts

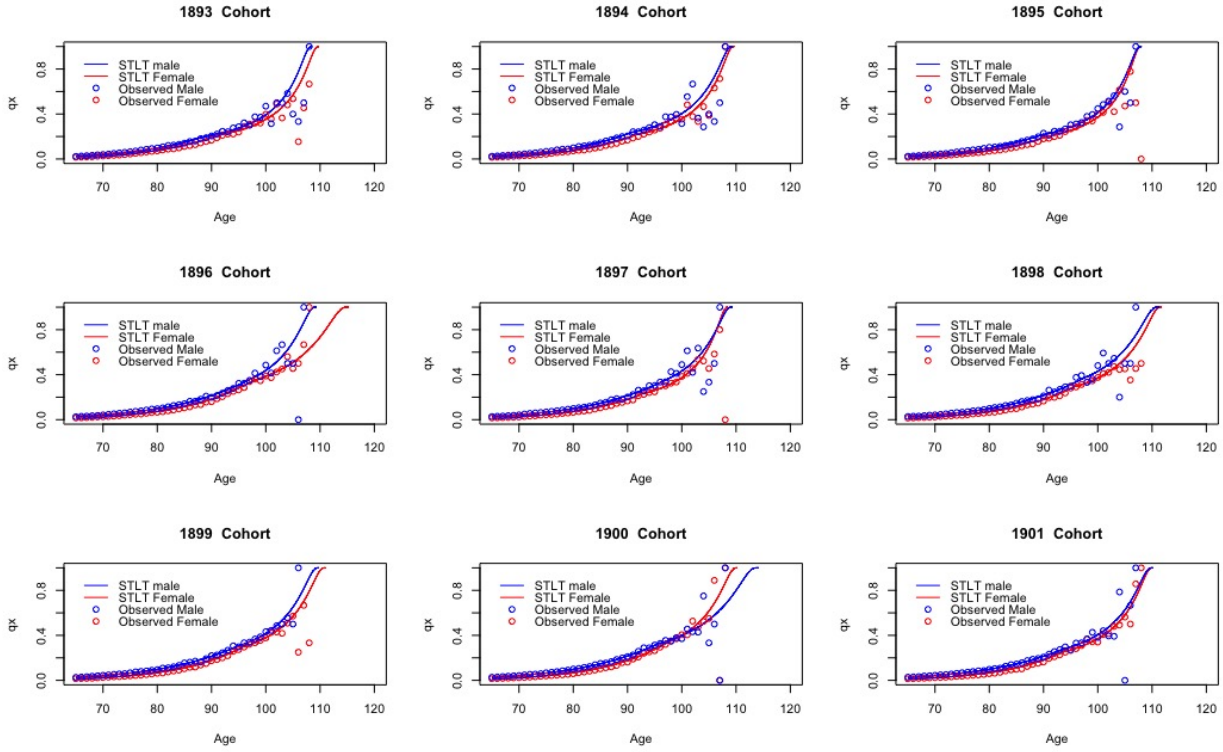


Figure 28: *Fitted STLT lines for all cohorts. Blue line corresponds to males; red line corresponds to females.*

References

Australian Government Actuary. (2019). *Construction of the Australian Life Tables*. http://www.aga.gov.au/publications/life_table_2010-12/06-Part2.asp [Accessed: 2019.08.10].

Balkema, A., & de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, *2*, 792–804.

Barbi, E., Lagona, F., Marsili, M., Vaupel, J. W., & Wachter, K. W. (2018). The plateau of human mortality: Demography of longevity pioneers. *Science*, *360*(6396), 1459–1461.

Basel Committee on Banking Supervision. (2016). *Minimum capital requirements for market risk*. Bank for International Settlements.

Bernard, J., & Vaupel, J. (1999). *Validation of exceptional longevity*. Odense University Press.

Boleslawski, L., & Tabeau, E. (2001). Comparing theoretical age patterns of mortality beyond the age of 80. In E. Tabeau, A. van den Berg Jeths, & C. Heathcote (Eds.), *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective* (pp. 127–155). Dordrecht: Springer Netherlands.

Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, *3*(1-2), 3–43.

Bourbeau, R., & Desjardins, B. (2002). Dealing with problems in data quality for the measurement of mortality at advanced ages in Canada. *North American Actuarial Journal*, *6*(3), 1–13.

Cairns, A., Blake, D., & Dowd, K. (2008). Modelling and management of mortality risk: A review. *SSRN Electronic Journal*.

Cairns, A., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., & Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, *48*(3), 355–367.

Chartres, B., & Stepleman, R. (1972). A general theory of convergence for numerical methods. *SIAM Journal on Numerical Analysis*, *9*(3), 476–492.

Coale, A., & Guo, G. (1989). Revised regional model life tables at very low mortality. *Population Index*, *55*, 614–43.

Coale, A., & Kisker, E. E. (1990). Defects in data on old-age mortality in the United States: New procedures for calculating mortality schedules and life tables at the highest ages. *Asian and Pacific Population Forum*, *4*, 1–31.

Currie, I. (2011). Modelling and forecasting the mortality of the very old. *ASTIN Bulletin*, *41*(2), 419–427.

Currie, I., Durban, M., & Eilers, P. H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, *4*(4), 279–298.

Dowd, K., & Blake, D. (2006). After VaR: The theory, estimation, and insurance applications of quantile-based risk measures. *Journal of Risk and Insurance*, *73*(2), 193–229.

Duffie, D., & Pan, J. (1997). An overview of value at risk. *Journal of Derivatives*, *4*(3), 7–49.

Dupuis, D. J., & Field, C. (1998). Robust estimation of extremes. *Canadian Journal of Statistics*, *26*, 199–215.

Einmahl, J. J., Einmahl, J. H. J., & de Haan, L. (2019). Limits to human life span through extreme value theory. *Journal of the American Statistical Association*.

Embrechts, P., Kluppelberg, C., & Mikosch, T. (2013). Modelling extremal events: For insurance and finance. *33*.

Embrechts, P., Resnick, S. I., & Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, *3*(2), 30–41.

Ferreira, A., & Huang, F. (2018). Is human life limited or unlimited? (a discussion of the paper by Holger Rootzén and Dmitrii Zholud). *Extremes*, *21*(3), 373–382.

Fisher, R., & Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*, 180–190.

Gavrilov, L., & Gavrilova, N. (1979). Determination of species length of life. *Doklady Biological Sciences*, *246*, 905–908.

Gavrilov, L., & Gavrilova, N. (1991). *The biology of life span : A quantitative approach*. Chur, Switzerland: CRC Press.

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

Gbari, S., Poulain, M., Dal, L., & Denuit, M. (2017). Extreme value analysis of mortality at the oldest ages: A case study based on individual ages at death. *North American Actuarial Journal*, *21*(3), 397–416.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality and on the mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society (London)*, *115*, 513–85.

Greenwood, M., & Irwin, J. O. (1939). The biostatistics of senility. *Human Biology*, *11*(1), 1–23.

Heligman, L., & Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, *107*(1), 437–55.

Human Mortality Database. (2019). University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). www.mortality.org [Accessed: 2019.07.01].

Institute and Faculty of Actuaries. (2016). *Solvency II - life insurance*. Institute and Faculty of Actuaries.

Jasilionis, D. (2018). *About mortality data for the Netherlands*. Human Mortality Database, University of California, Berkeley and Max Planck Institute for Demographic Research.

Kartashov, V., Maurer, R., Mitchell, O. S., & Rogalla, R. (2011). *Lifecycle portfolio choice with systematic longevity risk and variable investment-linked deferred annuities* (Working Paper No. 17505). National Bureau of Economic Research.

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

Li, J. S.-H., Hardy, M. R., & Tan, K. S. (2008). Threshold life tables and their applications. *North American Actuarial Journal*, *12*(2), 99–115.

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician*, *46*(1), 27–29.

Olshansky, S., & Carnes, B. (1997). Ever since Gompertz. *Demography*, *34*(1), 1–15.

Peng, L., & Welsh, A. (2001). Robust estimation of the generalized pareto distribution. *Extremes*, *4*(1), 53–65.

Purcal, S. (2006). *Supply challenges to the provision of annuities*. University of New South Wales, Working Paper.

Renshaw, A. E., & Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, *38*(3), 556–570.

Robine, J.-M., & Caselli, G. (2005). An unprecedented increase in the number of centenarians. *Genus*, *61*(1), 57–82.

Rootzén, H., & Zholud, D. (2017). Human life is unlimited – but short. *Extremes*, *20*(4), 713–728.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, *72*(1), 67–90.

Steenstrup, T., Kark, J. D., Verhulst, S., Thinggaard, M., Hjelmborg, J. V. B., Dalgard, C., ... Aviv, A. (2017). Telomeres and the natural lifespan limit in humans. *Aging*, *9*(4), 1130–1142.

Tan, P., & Drossos, C. (1975). Invariance properties of maximum likelihood estimators. *Mathematics Magazine*, *48*(1), 37–41.

Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society. Series A, 162*(1), 5–43.

Vaupel, J., & Robine, J.-M. (2002). Emergence of supercentenarians in low mortality countries. *North American Actuarial Journal*, *6*(3), 52–63.

Vincent, P. (1951). La mortalite des vieillards. *Population*, *6*(2), 181–204.

Watts, K. A., Dupuis, D. J., & Jones, B. L. (2006). An extreme value analysis of advanced age mortality data. *North American Actuarial Journal*, *10*(4), 162–178.

Wilmoth, J. (1995). Are mortality rates falling at extremely high ages: An investigation based on a model proposed by Coale and Kisker. *Population Studies*, *49*(2), 281–295.

Young, G. A. (2019). Mathematical statistics: An introduction to likelihood based inference. *International Statistical Review*, *87*(1), 178–179.