MortalityLaws: An R Package for Mortality Modelling

Marius Pascariu* and Vladimir Canudas-Romo

Max-Planck Institute for Biodemography of Aging,
University of Southern Denmark

March 21, 2017

Abstract

This paper provides an overview of the main mathematical models used by researchers over time, in mortality modelling. We want to outline a framework for developing forecasts of future mortality measures: age-specific death rates, death counts, mortality probabilities, and expectation of life, in order to better understand the mortality evolution. The existing laws of mortality can be assessed and investigated based on their general characteristics and ability to explain historical patterns of mortality. We also introduce an *R* package **MortalityLaws** which exploits the available optimization methods to provide tools for fitting a wide range of complex mortality models using HMD and assessing their goodness of fit.

Keywords: mortality law, density function, optimization, maximum likelihood, age patterns of mortality

1 Short history of mortality modeling

Modeling human mortality has been an important and active area of research for demographers, actuaries and medical scholars since Graunt (1662) first examined mortality in London to produce the first publication that was concerned mostly with public health statistics. Gaunts work showed that while individual life length was uncertain, there was a more predictable pattern of mortality in groups and causes of death. Halley (1693) showed how to actually construct a non-deficient mortality table from empirical birth-death data and even succeeded to present a method to perform a life annuity calculation based on this table. Such early tables were empirical and calculation was time consuming. Theoretical mortality modeling firstly began with DeMoivre (1725) who postulated a uniform distribution of deaths model, and showed simplified annuity calculation methods. Taking a biological approach to mathematical modeling, Gompertz (1825) assumed that force of mortality μ_x at age x in adulthood shows a nearly exponential increase,

$$\mu_x = \alpha e^{\beta x}, \tag{1}$$

where the two parameters α and β are positive; α varies with the level of mortality and β measures the rate of increase in mortality with age. The Gompertz model and its modified version by Makeham (1867), where an additional constant c is added to take into account the background mortality due to causes unrelated to age, were widely used as the standard models for adult mortality in humans (Kirkwood 2015, Olshansky & Carnes 1997); and then extended further to animal species in general (Sacher 1977),

$$\mu_x = \alpha e^{\beta x} + c. \tag{2}$$

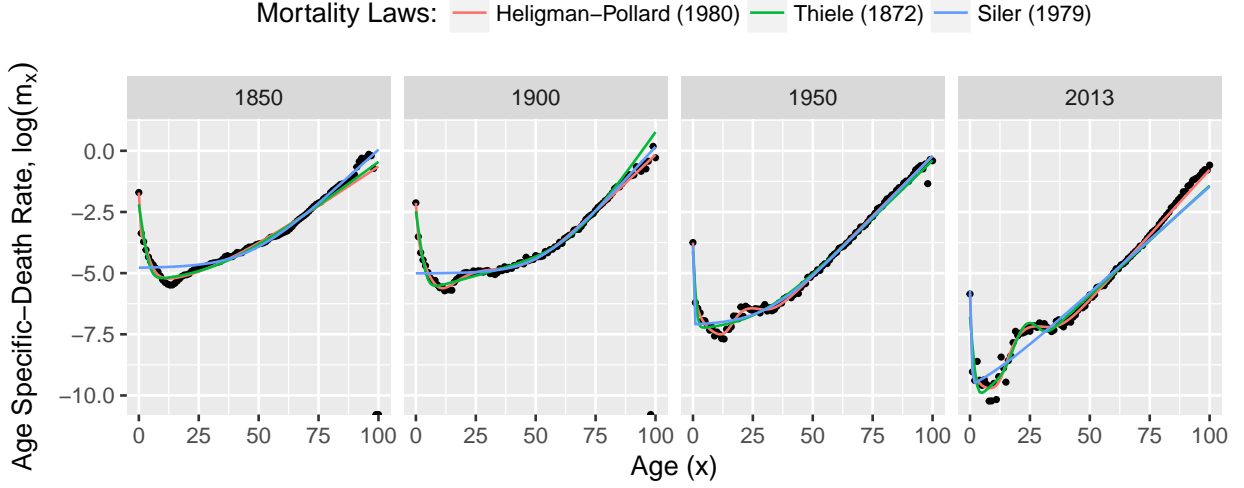*Corresponding author: mpascariu@health.sdu.dk

Figure 1: Observed and fitted death rates between age 0 and 80 for male population in Sweden. The mortality is extrapolated up to age 100.

Later on Heligman & Pollard (1980) proposed an eight-component mortality model,

$$\mu_x = A^{(x+B)^C} + De^{-E(\log x - \log F)^2} + GH^x. \tag{3}$$

Figure 1 shows how this model can fit the entire age-grange by decomposing the age pattern of mortality into three pieces. Each part with a relatively small number of parameters to control it. There are three parameters (A, B and C) to describe child mortality, three to describe a very flexible accident hump (D, E and F) typically occurring in young adulthood, and finally two parameters (G and H) to describe mortality at older ages. The main disadvantage of this model is that in its traditional form is difficult to fit and it does not account for uncertainty.

Siler (1983) developed a five-parameter competing hazard model in order to capture the mortality during "immaturity", adulthood and senescence and to facilitate inter-specific comparison,

$$\mu_x = a_1 e^{-b_1 x} + a_2 + a_3 e^{b_3 x}, \tag{4}$$

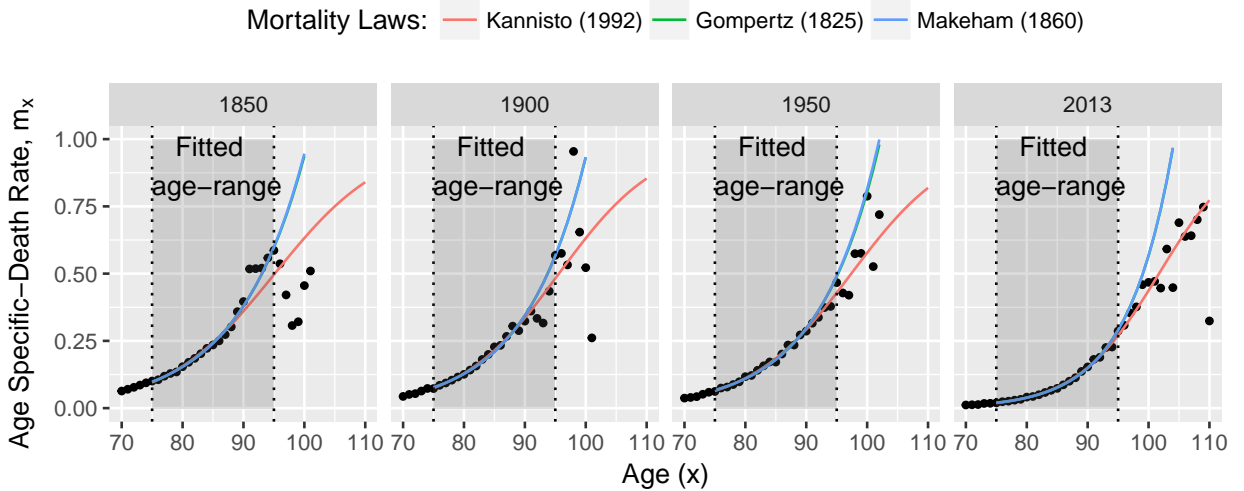where a_i, and b_i are location and dispersion parameters respectively.

Figure 2: Observed and fitted old-age mortality for female population in Sweden

2

Thatcher et al. (1998) performed studies to fit different mathematical models to different reliable data sets on adult and oldest-old mortality (aged 80 and above) covering the few recent decades. They evaluated the comparative compatibility of those models to the data, established the logistic model as the best mathematical model of human adult mortality, replacing the widely used Gompertz model and Makeham model. The logistic model assumes that the force of mortality μ_x is a logistic function of the age x.

$$\mu_x = \frac{\alpha e^{\beta x}}{1 + \alpha e^{\beta x}}, \tag{5}$$

the above models describe the mortality at a fix point in time; however actual mortality is stochastic and evolve continuously. Thus, while the mortality models described above are static, the parameters must be re-fit periodically to accommodate changes in mortality patterns.

Other important parametric functions or "laws" of mortality were developed by Thiele & Sprague (1871), Wittstein & Bumsted (1883), Steffensen (1930), Perks (1932), Harper (1936) and Weibull (1951).
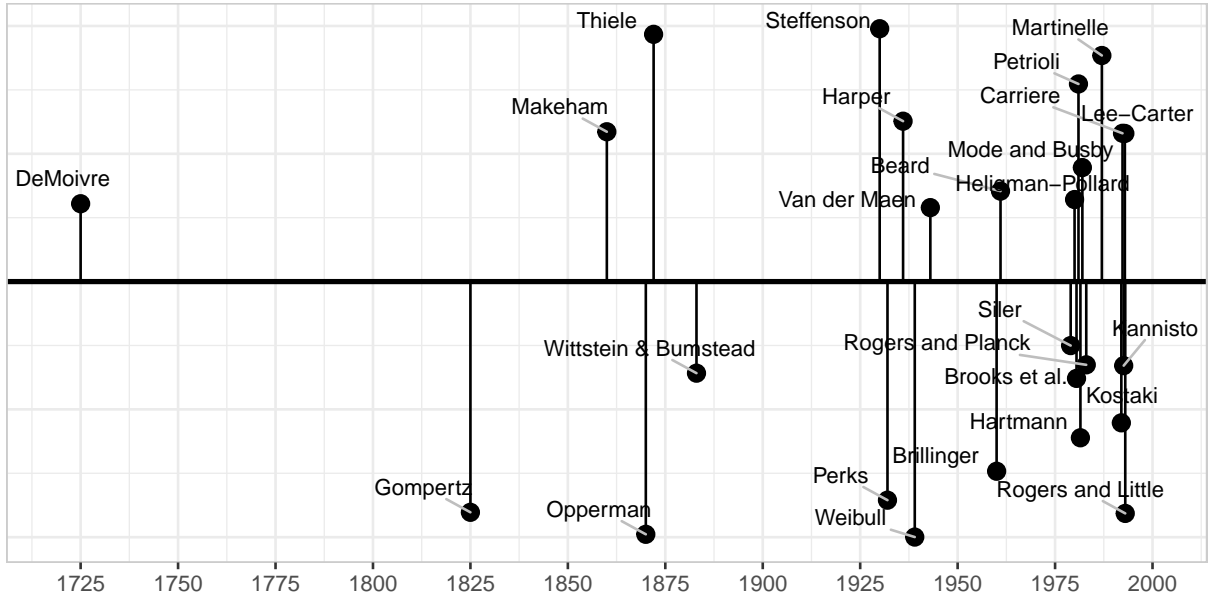

Figure 3: Mortality modelling timeline

3

Table 1: *Mortality laws* build in the **MortalityLaws** package

Mortality laws	Predictor	Code[†]
Gompertz	ae^{bx}	`gompertz0`
Gompertz	$\frac{1}{\sigma}exp\left\{\frac{x-m}{\sigma}\right\}$	`gompertz`
Inverse-Gompertz	$\frac{1}{\sigma}exp\left\{\frac{x-m}{\sigma}\right\}/\left(exp\left\{e^{\frac{-(x-m)}{\sigma}}\right\}-1\right)$	`invgompertz`
Makeham	$ae^{bx}+c$	`makeham0`
Makeham	$\frac{1}{\sigma}exp\left\{\frac{x-m}{\sigma}\right\}+c$	`makeham`
Inverse-Makeham	$\frac{1}{\sigma}exp\left\{\frac{x-m}{\sigma}\right\}/\left(exp\left\{e^{\frac{-(x-m)}{\sigma}}\right\}-1\right)+c$	`invmakeham`
Opperman	$\frac{a}{\sqrt{x}}+b+c\sqrt[3]{x}$	`opperman`
Thiele	$a_1e^{-b_1x}+\ a_2e^{-\frac{1}{2}b_2(x-c)^2}+a_3e^{b_3x}$	`thiele`
Wittstein & Bumstead	$\frac{1}{m}a^{-(mx)^n}+a^{-(M-x)^n}$	`wittstein`
Weibull	$\frac{1}{\sigma}\left(\frac{x}{m}\right)^{\frac{m}{\sigma}-1}$	`weibull`
Inverse-Weibull	$\frac{1}{\sigma}\left(\frac{x}{m}\right)^{-\frac{m}{\sigma}-1}/\left(exp\left\{\left(\frac{x}{m}\right)^{-\frac{m}{\sigma}}\right\}-1\right)$	`invweibull`
Siler	$a_1e^{-b_1t}+a_2+a_3e^{b_3t}$	`siler`
Heligman - Pollard	$A^{(x+B)^C}+De^{-E(\ln x\ -\ln F\ )^2}+GH^x$	`HP`
Kannisto	$\frac{ae^{bx}}{1+ae^{bx}}+c$	`kannisto`
Carriere[*]	$s\left(x\right)=\psi_1s_1\left(x\right)+\ \psi_2s_2\left(x\right)+\ \psi_3s_3\left(x\right)$	`carriere1`

[†] Code to be used in **MortalityLaw** function in order to specify the model
[*] In Carriere different survival function are combined in order to obtain the right pattern of mortality (for example Carriere1 = Weibull + Inverse Weibull + Gompertz). $\psi_{1,2,3}$ are weighting positive factors where $\psi_1+\psi_2+\psi_3=1$.

2   MortalityLaws R package

MortalityLaws in an R package which exploits the available optimization methods to provide tools for fitting and analyzing a wide range of complex mortality models. Three main functions are available in the package: `ReadHMD`, `MortalityLaw` and `lifetable`. The package provides also generic functions like `summary`, `plot`, `coef`, `fitted.values` and a small data set for testing purposes `HMD.test.data`.
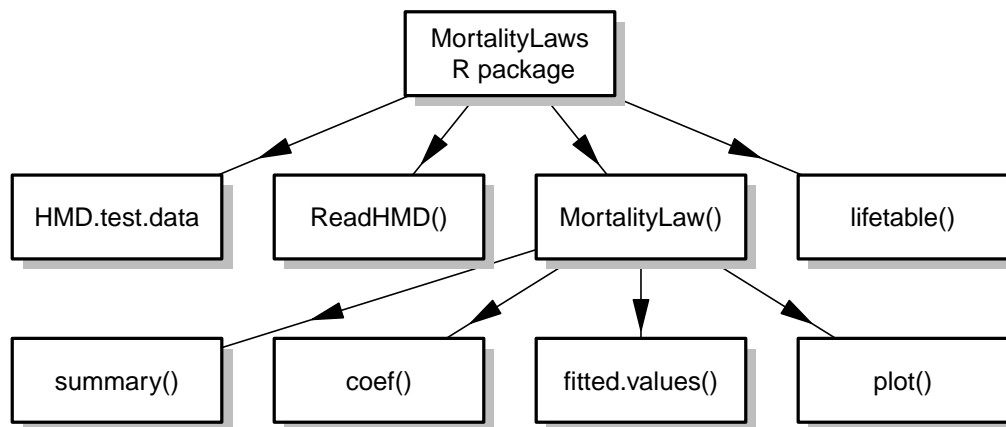

Figure 4: The structure of the MoralityLaws R package

3   Fitting mortality models with MortalityLaws

In this section we describe how to use the **MortalityLaws** package for fitting human mortality data.

3.1   Installation

MortalityLaws is a software in development. The repository containing the development version of the package can be found on GitHub and can be installed by running the following code in the R console:

```
# install.packages("devtools")
library(devtools)
install_github("mpascariu/MortalityLaws")
```

The package is loaded within **R** as follows:

```
library(MortalityLaws)
```

3.2 Data

Download data form Human Mortality Database (2016) using the `ReadHMD` function:

```r
country = "SWE" # HMD country code for Sweden

# Download HMD data - death counts
HMD_Dx <- ReadHMD(what = "Dx",
                  countries = country,
                  interval = "1x1",
                  username = "user@email.com", # here add your HMD username
                  password = "password", # here add your password account
                  save = FALSE)

# In the same way we download the exposures
HMD_Ex <- ReadHMD(what = "Ex", countries = country, interval = "1x1",
                  username = "user@email.com", password = "password",
                  save = FALSE)
```

The downloaded object should look like this:

```
HMD_Dx

## Human Mortality Database (www.mortality.org)
## Downloaded by: pascariu.marius@outlook.com
## Download Date: Tue Mar 21 12:51:58 2017
## Type of data: Dx
## Countries included: SWE
##
## Data:
##         country Year Age  Female    Male   Total
## 1           SWE 1751   0    5988    6902   12890
## 2           SWE 1751   1 1286.45 1359.88 2646.33
## 3           SWE 1751   2  834.56  882.13 1716.69
## 4           SWE 1751   3     622  655.26 1277.26
## 5           SWE 1751   4  470.99  497.74  968.73
## ...        <NA>  ... ...     ...     ...     ...
## 29300       SWE 2014 106   10.78    0.96   11.74
## 29301       SWE 2014 107    5.08       0    5.08
## 29302       SWE 2014 108    2.62       0    2.62
## 29303       SWE 2014 109    1.18       0    1.18
## 29304       SWE 2014 110    0.31       0    0.31
```

Here we downloaded all the registered death counts in Sweden from 1751 until 2014. In the same way one can download the following records: birth counts, population size, lexis triangles, exposures, life tables, life expectacy at birth and death rates for over 38 countries and regions in different formats.

3.3 Model fitting and diagnosis

Once we have data from HMD or other sources we can start analyzing it. For example, let's fit a Heligman & Pollard (1980) model under a Poisson setting which is already implemented as one of the standard models in the package[1]. We have to use the `MortalityLaw` function in this regard.

---

[1] Check the Appendix to see all the implemented mortality laws.

```r
# Select 1 year of data, an age-range and sex
year = 1975
ages = 0:100
sex = "Male" # in HMD we have 'Female', 'Male', 'Total'
Dx <- HMD_Dx$data
Ex <- HMD_Ex$data

x = ages - min(ages) # scale ages in order to get meaningful parameter estimates
deaths <- Dx[Dx$Year == year & Dx$Age %in% ages, ]
exposure <- Ex[Ex$Year == year & Ex$Age %in% ages, ]

fit <- MortalityLaw(x = x,
                    Dx = deaths[, sex], # vector with death counts
                    Ex = exposure[, sex], # vector containing exposures
                    law = "HP",
                    how = "poissonL",
                    fit.this.x = 0:65)

ls(fit) # inspect the output object

## [1] "coefficients"        "fitted.values"       "goodness.of.fit"
## [4] "info"                "input"               "optimization.object"
## [7] "residuals"
```

A summary can be obtained using the `summary` function.

```r
summary(fit)

## Heligman-Pollard (1980):
## q(x)/p(x) = a^((x+b)^c) + d*exp(-e*(log(x/f))^2) + g*h^x
##
## Deviance Residuals:
##     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.19400  -0.00020   0.00019  -0.00333   0.00140   0.04461
##
## Coefficients:
##        a        b        c        d        e        f        g        h
##  0.00050  0.00400  0.08000  0.00100  9.99999 17.00030  0.00005  1.09945
##
## Log-Likelihood =  11.33  AIC =  -6.66  BIC =  10.86
```

The standard plot helps us to investigate visually the goodness of fit.
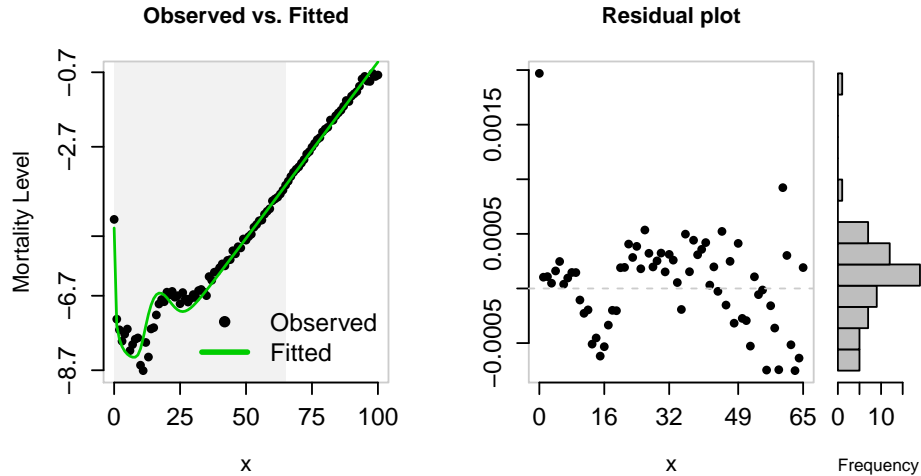
```r
plot(fit)
```

Figure 5: Heligman-Pollard model fitted using the MortalityLaw function

The gray area on the plot showing the fitted value indicates the age range used in fitting the model. This is specified in the model design with the help of `fit.this.x` argument, which can be adjusted accordingly.

We can use the fitted values to compute life tables. This can be achieved using the `lifetable` function.

```
lt <- lifetable(x = ages, mx = fitted.values(fit))
ls(lt)

## [1] "lt"          "lt.exact"      "process_date"

head(lt$lt)

##   age       mx       qx  ax     lx  dx    Lx      Tx    ex
## 1   0 0.007595 0.007544 0.1 100000 754 99925 7273150 72.79
## 2   1 0.000554 0.000554 0.5  99246  55 99218 7173225 72.30
## 3   2 0.000384 0.000384 0.5  99191  38 99172 7074007 71.33
## 4   3 0.000315 0.000315 0.5  99153  31 99137 6974835 70.36
## 5   4 0.000278 0.000278 0.5  99121  28 99108 6875698 69.38
## 6   5 0.000256 0.000256 0.5  99094  25 99081 6776591 68.39
```

Now let's fit a mortality law that is not defined in the package, say a reparametrize version of Gompertz in terms of modal age at death (Missov et al. 2015),

$$\mu_x = \beta e^{\beta(x-M)}. \tag{6}$$

We have to define a function containing the desired hazard function and then using the `custom.law` argument it can be used in the `MortalityLaw` function.

```
# Select the data that we want to fit
year = 2014
ages = 25:75
sex = "Female"
deaths <- Dx[Dx$Year == year & Dx$Age %in% ages, ]
exposure <- Ex[Ex$Year == year & Ex$Age %in% ages, ]
```

```r
# Here we define a function for our new model and provide start parameters
my_gompertz <- function(x, par = c(b = 0.13, M = 45)){
  hx  <- with(as.list(par), b*exp(b*(x - M)) )
  return(as.list(environment())) # return everything inside this function
}

# Use 'custom.law' argument to instruct the MortalityLaw function how to behave
my_model <- MortalityLaw(x = ages,
                         Dx = deaths[, sex],
                         Ex = exposure[, sex],
                         custom.law = my_gompertz)
summary(my_model) # these are the results

## Custom Model
##
## Deviance Residuals:
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.00049 -0.00012  0.00003  0.00001  0.00014  0.00048
##
## Coefficients:
##         b        M
##   0.09703 91.91063
##
## Log-Likelihood =  11.07  AIC =  -18.15  BIC =  -14.28

plot(my_model)
```
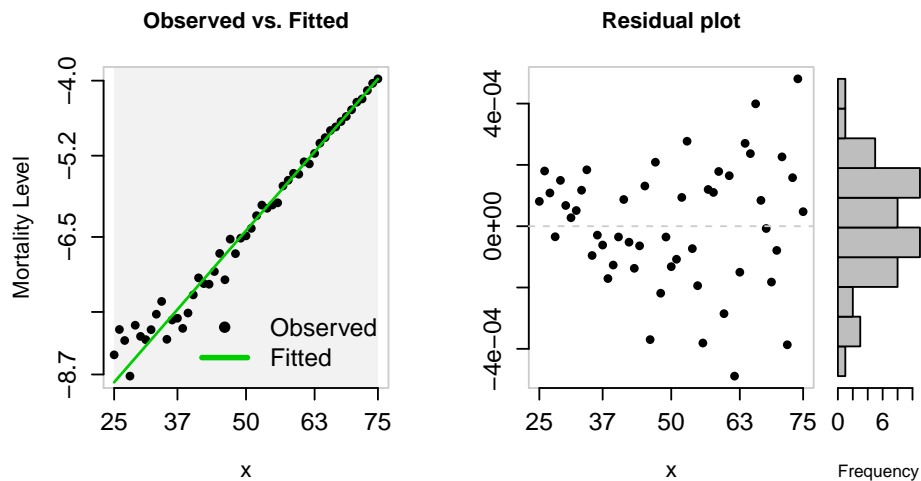

Figure 6: Gompertz model fitted using the MortalityLaw function

References

DeMoivre, A. (1725), 'Annuities on lives: Or, the valuation of annuities upon any number of lives as also of reversions', *William Person, London* .

Gompertz, B. (1825), 'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies', *Philosophical transactions of the Royal Society of London* **115**, 513–583.

Graunt, J. (1662), 'Natural and political observations mentioned in a following index, and made upon the bills of mortality.', *London: John Martyn and James Allestry, 1973* .

Halley, E. (1693), 'An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of breslaw; with an attempt to ascertain the price of annuities upon lives, pt nr. 196'.

Harper, F. S. (1936), 'An actuarial study of infant mortality', *Scandinavian Actuarial Journal* **1936**(3-4), 234–270.

Heligman, L. & Pollard, J. H. (1980), 'The age pattern of mortality', *Journal of the Institute of Actuaries* **107**(01), 49–80.

Kirkwood, T. B. (2015), 'Deciphering death: a commentary on Gompertz (1825)'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies", *Phil. Trans. R. Soc. B* **370**(1666), 20140379.

Makeham, W. M. (1867), 'On the law of mortality', *Journal of the Institute of Actuaries (1866-1867)* **13**(6), 325–358.

Missov, T. I., Lenart, A., Nemeth, L., Canudas-Romo, V. & Vaupel, J. W. (2015), 'The gompertz force of mortality in terms of the modal age at death', *Demographic Research* **32**, 1031–1048.

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

Perks, W. (1932), 'On some experiments in the graduation of mortality statistics', *Journal of the Institute of Actuaries (1886-1994)* **63**(1), 12–57.

Sacher, G. A. (1977), 'Life table modification and life prolongation', *Handbook of the Biology of Aging* **582**.

Siler, W. (1983), 'Parameters of mortality in human populations with widely varying life spans', *Statistics in medicine* **2**(3), 373–380.

Steffensen, J. (1930), 'Om sandsynligheden for at afkommet uddør', *Matematisk tidsskrift. B* pp. 19–23.

Thatcher, A. R., Kannisto, V. & Vaupel, J. W. (1998), 'The force of mortality at ages 80 to 120.'.

Thiele, T. N. & Sprague, T. (1871), 'On a mathematical formula to express the rate of mortality throughout the whole of life, tested by a series of observations made use of by the danish life insurance company of 1871', *Journal of the Institute of Actuaries and Assurance Magazine* **16**(5), 313–329.

University of California Berkeley, USA & Max Planck Institute for Demographic Research, Germany (2016), 'Human Mortality Database'.
URL: *http://www.mortality.org/*

Weibull, W. (1951), 'Wide applicability', *Journal of applied mechanics* **103**, 293–297.

Wittstein, T. & Bumsted, D. (1883), 'The mathematical law of mortality', *Journal of the Institute of Actuaries and Assurance Magazine* **24**(3), 153–173.

4 Appendix

Table 2: Loss functions

Name	Function	Code[†]
Poisson Log-Likelihood	$\sum_x \left\{ D_x \log \widehat{\mu}_x \ - \ E_x^c \widehat{\mu}_x \right\} + c$	`poissonL`
Binomial Log-Likelihood	$\sum_x \left\{ D_x \log \left[1 - e^{-\widehat{\mu}_x} \right] \ - \ [E_x^c - D_x] \widehat{\mu}_x \right\} + c$	`binomialL`
Loss Function 1	$\sum_x \left(1 - \frac{\widehat{\mu}_x}{\mu_x} \right)^2$	`LF1`
Loss Function 2	$\sum_x \log \left(\frac{\widehat{\mu}_x}{\mu_x} \right)^2$	`LF2`
Loss Function 3	$\sum_x \frac{(\mu_x - \widehat{\mu}_x)^2}{\mu_x}$	`LF3`
Loss Function 4	$\sum_x (\mu_x - \widehat{\mu}_x)^2$	`LF4`
Loss Function 5	$\sum_x (\mu_x - \widehat{\mu}_x) \log \left(\frac{\mu_x}{\widehat{\mu}_x} \right)$	`LF5`
Loss Function 6	$\sum_x |\mu_x - \widehat{\mu}_x|$	`LF6`

[†] Code to be used in **MortalityLaw** function in order to specify the fitting procedure.