A Modeling Approach to Life Table Data Sets

Christos H. Skiadas[1] and Charilaos Skiadas[2]

[1] Technical University of Crete
University Campus
73100 Chania, Crete, Greece
(e-mail: skiadas@ermes.tuc.gr)

[2] Hanover College
Hanover Indiana, USA
(e-mail: skiadas@hanover.edu)

Abstract. A modeling approach to Life Table Data sets is proposed. The method is based on a stochastic methodology and the derived first exit time probability density function. The Health State Function of a population is modeled as the mean value of the health states of the individuals. The form for the health state function suggested here is relatively simple compared to previous attempts but the application results are quite promising. The model proposed is a three parameter model and it is compared to the three parameter Weibull model. Both models are applied to the Life Table Data for males and females in Greece from 1992 to 2000. The results indicate that the proposed model fits better to the data than the Weibull model. The methodology for the model building and the model proposed could be used in several cases in population studies in biology, ecology and in other fields.
Keywords: Life Table Data, Health State Function, Stochastic modelling, Hitting time density function, Gompertz model, Weibull model.

1 Introduction

Many attempts were done during the last centuries to model Life Table Data and the inevitable decay process of an original population during time. The most important for this study are the models proposed by Gompertz to express the law of human mortality and the Weibull model as a model to express the failure of items in a set of products. The later is also a flexible model to express the distribution function of the number of deaths of a population. The task of this study is to propose a simple three parameter model that can be applied to life table data based on a stochastic theory presented in previous studies. In these studies more complicated models where proposed and applied. However, these models where quite heavy to handle and apply. Furthermore, as these models have several parameters, it is not possible to test their fitting ability relative to the simpler three parameter models in use, like for instance the Gompertz the Weibull models.

2 Model Analysis

The *Gompertz model*, proposed by Benjamin Gompertz (1825) and analysed by other researchers many years ago for the analysis needed in this study (Winsor, 1932), has the form of the following density function:

$$g(t) = ke^{-\ell t} \exp\left(-be^{-\ell t}\right)$$

The model is left skewed and it is not easy to apply to life table data in all the data range. Instead it is a good model to express the mortality data from an age close to 30 years up to a maximum level of the death rate, usually between 70-80 years depending on the data set used (males or females), (Haybittle, 1998).

The Weibull model was proposed by Waloddi Weibull (1951). The probability density function of this model has the form:

$$f(x;k,\lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^{k}}$$

Usually this is considered as a 2-parameter model. However, when applying the model to data as is the case in this study an extra parameter is present. Then the model takes the following 3-parameter form:

$$f(x;k,\lambda,c) = c\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^{k}}$$

where the parameter c will be determined when fitting the model to the data.

3 The Proposed Model

In previous studies a general model was proposed, based on the theory of the hitting time of a stochastic process in a barrier located at a distance a from the horizontal axis which expresses the age of the individual (Janssen and Skiadas, 1995; Skiadas, 2006b,a). The stochastic variable expresses the health state of the individual whereas the mean value of the process expresses the health state of the population.

This model, for the case of the state of human health, has the following probability density function (Skiadas, 2006b,a):

$$g(t) = k^* \frac{|\alpha^*|}{\sigma\sqrt{2\pi t^3}} \exp\left(-\frac{\left(\alpha^* - H(t)\right)^2}{2\sigma^2 t}\right) \quad (1)$$

where $k*$ is a normalisation constant defined by the formula:

$$\int_0^\infty g(t)\,dt = 1$$

Where $H(t)$ is the *health state function* and σ is the variance.

Although the concept of the health and the health state are terms very much defined and used in our societies from the very beginning, a mathematical analytic form is not obtained. It is a common knowledge that the health state decrease over time, and also that frequent and sudden changes appear, thus leading to the acceptance that the health state follows a stochastic path during time.

Regarding the mean value of the health of a population denoted by $H(t)$ in equation (1), this must be expressed by a function of unknown form which is decreasing at least for large values of the age t. The development and application of the Gompertzian theory in the last two centuries indicated that a rapidly decreasing function may express the state of human health in the last years of the human life very well (Haybittle, 1998).

Instead, in the middle period of human life, the mean health state could be expressed by a slowly decreasing function of time. There remain the first years after birth, when the state of the human health is in relatively lower levels.

The modeling of this early period, along with the remaining life time, was done in previous studies by introducing a quite heavy model for the health state function. The application of this model to life table data was successful. However, it was not possible to compare it to simpler models, such as the Gompertz and Weibull models, due to the larger number of parameters of the proposed model. Also the proposed model was modeling the total life period including the first stages of human life, something that is not possible by applying the simpler models. Application of this model shows that the form of the health state function for large ages is mainly *flat* and slowly *decreasing*.

Another very important point is that when we expand the unknown health state function $H(t)$ in a Taylor series, the rapidly decreasing part of the function could be expressed simply by a term having a large exponent. This means that a function of the following form could be a simple but quite good approximation to the real situation:

$$H(t) = c - (\ell t)^b$$

where c, ℓ and b are positive parameters.

For $b > 0$ we see that $H(t)$ is a decreasing function. Especially for $b \gg 1$, then $H(t)$ is a rapidly decreasing function. Figure 1 illustrates this case for various values of the parameter b ($c = 20$, $\ell = 0.03$ and $b = 3, 4, 5$).

Introducing the above value for $H(t)$ in equation (1), the following form results:

$$g(t) = k(\ell t)^{-\frac{3}{2}} \exp\left(-\frac{\left(\alpha - (\ell t)^b\right)^2}{2t}\right) \tag{2}$$

where $a = a^* + c$, k is the new integration constant and the variance σ is included in the parameters k, a and ℓ. Without loss of generality, in many applications it could be assumed that $\sigma = 1$.

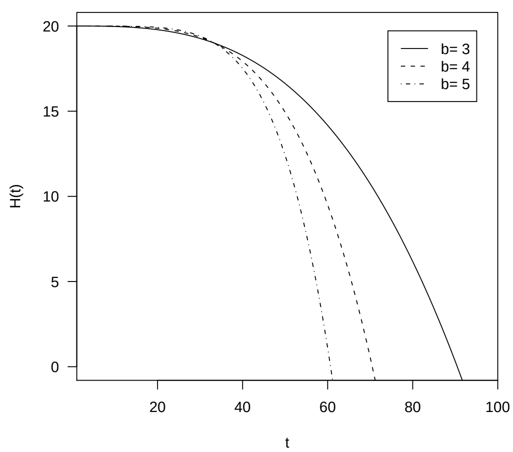

Fig. 1. Health state functions of the form $H(t) = c - (\ell x)^b$.

Figure 2 illustrates the above case for various values of the parameter b ($c = 20$, $\ell = 0.03$ and $b = 3, 4, 5$). It is clear that the higher the value of the parameter b, the faster the rate of decrease is, and the sharper the density function becomes.

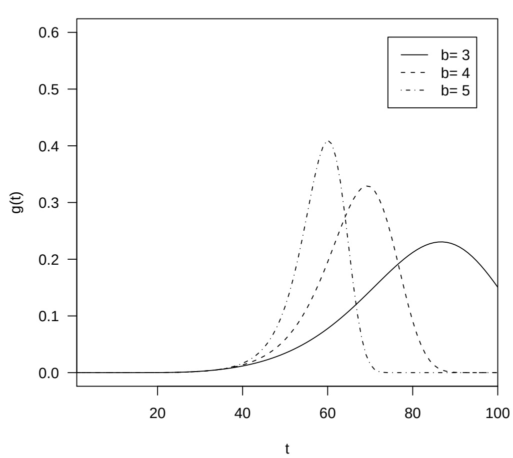

Fig. 2. The density function for a health state function of the form $H(t) = c - (\ell x)^b$.

4 Applications to Life Table Data and Comparisons

Depending on whether the life table data we are looking at are for male or female populations, a different value of b will be used ($b = 4$ for males and $b = 5$ for females is selected). The following table shows the results of fitting the life table data for Greece from 1992 to 2000 to our model, using a non-linear least squares fit, and comparing the fit with the corresponding fit from the Weibull distribution.

For the purposes of the performing fit, the parameters are set in a slightly different way. Here is the density function from the proposed model:

$$g(t; k, \ell, \alpha, b) = k\,(\ell t)^{-3/2} \exp\left(-\frac{\left(\alpha - (\ell t)^b\right)^2}{2t}\right)$$

And the corresponding equations for the Weibull model:

$$g(t; c, B, k) = c\,(tB)^{k-1} \exp\left(-(tB)^k\right)$$

Table 1 summarises the non-linear regression results for Greece (males) from 1992 to 2000. The proposed monomial model shows better fitting behavior compared to the Weibull model for all the nine years studied. Table 2 shows the corresponding results for females.

Another interesting point is to find out at which time T the value of the health state function $H(t)$ becomes zero. This is achieved when

$$T = \frac{\alpha^{1/b}}{\ell}$$

For the case of males ($b = 4$) the estimated value is $T = 82.1707$ in 1992 and $T = 82.69$ in the year 2000. There appears to be an increase of $t = 0.51929$, or approximately half of a year, in a 9 years period.

For the case of females the estimated value is $T = 85.37205$ in 1992 and $T = 86.61715$ in the year 2000. There appears to be an increase of $t = 1.2451$ in a 9 years period.

Using our estimates for the parameters of the model, we may construct a graph to illustrate the health state function for males and females at a specific year. The Figure 3 shows the health state function for males and females at the year 2000. As expected from the above theory, in both cases there appears to be a relatively stable period, represented by the flat part of the curve for the years up to 40 for females and up to 30 for males. Then a gradually decreasing period follows. The health state function shows higher values for females compared to that for males.

Figures 4 and 5 illustrate the raw data for Greece 2000 for females and males respectively, along with curves expressing the proposed model and the Weibull model. In both cases the fitting is quite good. The relatively higher sum of squared error for males is due to the sharp form of the data in the

	Proposed Model Fit				Weibull Fit			
Year	k	ℓ	α	SS	c	k	B	SS
1992	9.49813	0.02523	18.47361	3.95820	8.66475	7.41596	0.01236	5.26180
1993	9.51720	0.02523	18.47850	3.91172	8.68521	7.42218	0.01235	5.18750
1994	9.37214	0.02513	18.34295	4.00749	8.56078	7.30223	0.01234	5.13369
1995	9.35083	0.02508	18.33543	4.61074	8.54554	7.29311	0.01232	5.76867
1996	9.20313	0.02500	18.15040	3.74582	8.42971	7.15397	0.01232	4.62984
1997	9.25545	0.02503	18.19395	3.48055	8.47724	7.18893	0.01233	4.24332
1998	9.20220	0.02494	18.13499	3.75199	8.42760	7.11416	0.01230	4.28701
1999	9.26440	0.02500	18.20473	3.27281	8.47366	7.15850	0.01231	3.64226
2000	9.33965	0.02501	18.29502	3.72047	8.52884	7.20542	0.01230	3.95039

Table 1. Fit comparison for Greece, Males.

	Proposed Model Fit				Weibull Fit			
Year	k	ℓ	α	SS	c	k	B	SS
1992	10.50017	0.02149	20.76597	3.09603	11.42873	10.36549	0.01181	4.30740
1993	10.54319	0.02148	20.82900	3.49145	11.46462	10.40470	0.01180	4.80948
1994	10.64504	0.02145	21.02880	3.31293	11.55272	10.53462	0.01175	4.66292
1995	10.54154	0.02139	20.88549	2.92138	11.44183	10.38639	0.01174	4.20394
1996	10.48414	0.02135	20.82431	2.59704	11.38510	10.32809	0.01173	3.88427
1997	10.40077	0.02127	20.73785	2.62815	11.30523	10.24637	0.01170	4.00420
1998	10.42762	0.02123	20.80106	2.91107	11.32323	10.26573	0.01167	4.22223
1999	10.33880	0.02118	20.66995	2.00799	11.23670	10.14668	0.01166	3.22056
2000	10.35440	0.02115	20.61470	1.88820	11.24338	10.05560	0.01165	2.76724

Table 2. Fit comparison for Greece, Females.

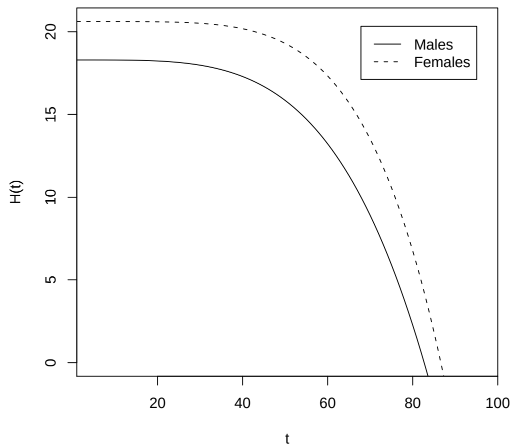

Fig. 3. Estimated health state functions for males and females, Greece 2000.

range of the maximum death rate, as is illustrated in Figure 5. For males the main part of the error term is due to fluctuations around the range of the maximum death, as well as the deaths at ages 18-28 years, which are mostly due to accidents and other reasons.

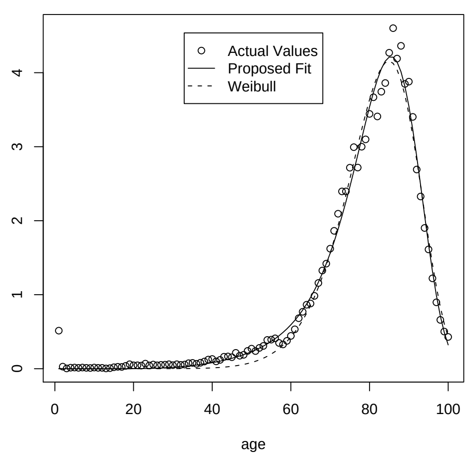

Fig. 4. Fit, Females, Greece, 2000

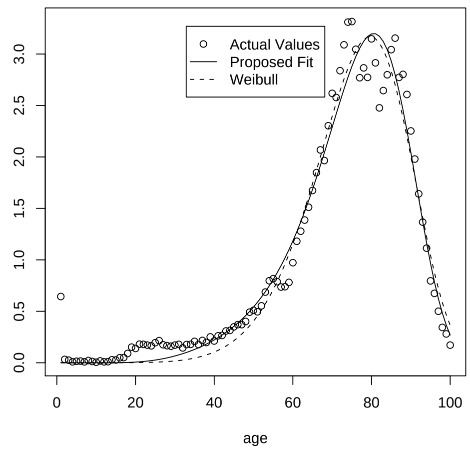

Fig. 5. Fit, Males, Greece, 2000

A comparative view of the sum of square residuals (SSR) is illustrated in the Figures 7 and 6, where the SSR of the proposed model are plotted against the SSR of the Weibull model. The results show an almost stable superiority of the proposed model for the case of females. In the males case the proposed model shows lower SSR in all the tested time period. However, the difference tends to shorten over time.

The results regarding the SSR of the proposed model show that perhaps a model with a varying parameter b could give better results. Also we can test our assumption regarding the values of the parameter b selected for males ($b = 4$) and females ($b = 5$). The parameter is close or higher to $b = 5$ for females whereas for males there appear values larger or smaller than $b = 4$. However this value is close to $b = 4$ for the years 1996 to 2000.

5 Conclusions

In this paper we proposed and applied a 3-parameter model to express the distribution function of the number of deaths of a population. This model was applied to the Greek life table data from 1992 to 2000. The proposed model was tested and compared with the Weibull model. The comparative results were quite promising, indicating a better fit for the proposed model than the commonly used Weibull model.

The good application results of the proposed model strengthen the underlying theoretical assumptions of the stochastic theory used. The modeling of life table data sets based on the hitting time theory and the resulting

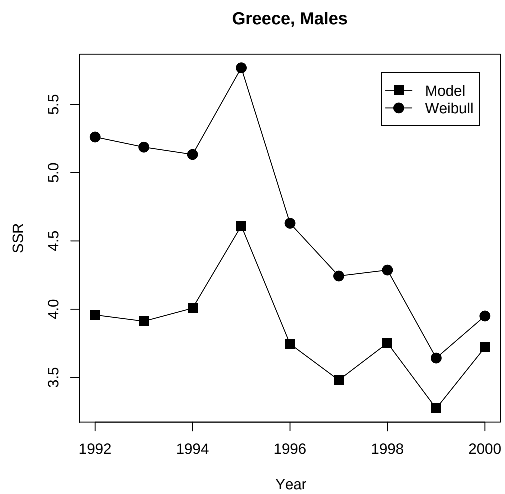

Fig. 6. Comparison of the sum of square residuals for the two models, Males.

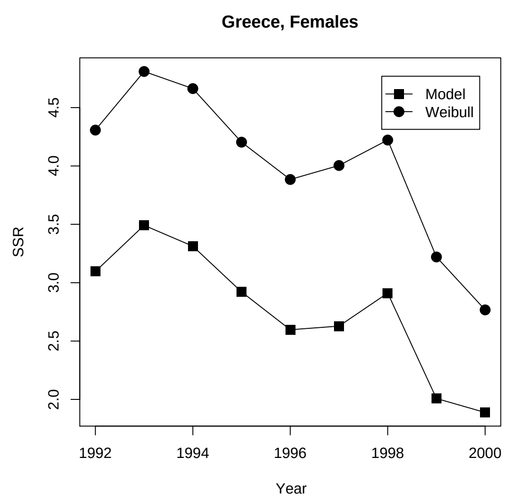

Fig. 7. Comparison of the sum of square residuals for the two models, Females.

probability density function seems to be a quite promising new direction of research.

Bibliography

B. Gompertz. On the nature of the function expressing of the law of human mortality. *Philosophical Transactions of the Royal Society*, 36:513–585, 1825.

J. L. Haybittle. The use of the gompertz function to relate changes in life expectancy to the standardized mortality ratio. *International Journal of Epidemiology*, 27:885–889, 1998.

J. Janssen and C. H. Skiadas. Dynamic modelling of life-table data. *Applied Stochastic Models and Data Analysis*, 11(1):35–49, 1995.

C. H. Skiadas. Dynamic modeling of greek life table data. In F. Vonta, M. Nikulin, N. Limnios, and C. Huber-Carol, editors, *Statistical Methods for Biomedical and Technical Systems*, pages 449–454. Birkhäuser, 2006.

C. H. Skiadas. Stochastic modeling of greek life table data. *Communications in Dependability and Quality Management*, 2006.

W. Weibull. A statistical distribution function with wide applicability. *Appl. Mech.*, 18:293–297, 1951.

C. P. Winsor. The Gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences*, 18(1):1–7, 1932.