MODELLING THE RISK OF MORTALITY IN ROMANIA [1]

Tiberiu DIACONESCU
Institutul de Prognoză Economică, Academia Română,
diaconescutiberiu@gmail.com

Abstract

In the last 3 centuries, researchers from different area of expertise, such as, demographers, medical doctors and actuarial mathematicians, have been struggling to develop a better model to estimate biometric functions. Such a model is long due in order to improve the present methodology regarding certain statistic-demographic rates. The issue of missing data (for older ages), the issue of computing correctly the average expected life and last but not least the forecast of mortality, could be solved through the use of better models that can determine the components of life tables. A life table represents a means of determining the probabilities of an individual living to or dying at a certain age. A better image of the ageing process for human population is shown by determining the probability of death occurring at specific ages over specified periods of time. Parametric models for the projection of mortality rates were first introduced by Lee and Carter (1992) in the US, representing an important development in demography. The model was followed by several others models which were developed over the years [Gompertz, 1825; Makeham, 1860; Weibull, 1951; Beard, 1959; Vaupel et al, 1979, Kannisto, 1992]. The present study is trying to provide certain information over the best use of these types of models under specific hypothesis. We will focus on the methodology of estimating the parameters for Gompertz' law of mortality and how well it can be fitted using data from Romania in 2012.

Keywords: *mortality risk, probability of death, mortality rate, laws of mortality, modeling risk of death*

---

[1] This paper represents a selected fraction from the doctoral thesis entitled: "Economic impact of the ageing population in the European Union", coordinator Acad. Lucian Liviu Albu., National Institute of Economic Research "Costin C. Kiritescu" , Romanian Academy, Bucharest.

Introduction

Since the first half of the 18th century, laws of mortality (parametric functions) that can be used to model empirical mortality curves, have developed into one of the most important work for demographers and actuarial mathematicians, but also to all others interested in the statistical study of human mortality.

One of the most well-known model, which can be interpreted as a parametric mathematical law, was proposed by Gompertz (1825), where the risk of mortality could be estimated by: $\mu(x)=Bc^x$, which in fact represents an exponential function. Both B and c are considered non negative and x represents the age of an individual, which most commonly is in an interval between 0 and 100. Most modern data are available only until the age of 84, but further complex studies use data well over the age of 100 years (according to the age of the elderly individuals) [Kannisto, 1994, Jeune and Vaupel, 1995, Kannisto, 1996].

In literature we can find also three laws of mortality that apply to all ages. Two of them were developed by Thiele and Wittstein in the late 19th century, the third, developed by Heligman and Pollard, more recently.

Gompertz modeled his law by studying the survival curves from life tables which were available at that time. He described it as a hypothesis and considered the consequences of its use at larger age intervals, though not including infancy or very old age mortality.

Later on, Brillinger (1961), said that if the human body was to be considered as a series system of independent components, then the force of mortality may follow Gompertz's law.

In the past, analytical approaches (such as the Gompertz'or Makeham's law) managed to satisfy this hypothesis approximately over a broad range of ages. However, as modern data have become more available and reliable, the uses of approximate have become less acceptable.

Nowadays, mortality is most commonly represented in the form of a life table, which gives probabilities of death or survival, within one year, at exact integral age. These probabilities are generally based on tabulations of deaths in a given population and estimates of the size of that population. Functions in the life table can be generated from q_x where q_x is the probability of death within a year of a person aged x.

Although a life table does not give mortality at non-integral ages or for non-integral durations, as can be obtained from a mathematical formula, acceptable methods for estimating such values are well known.

In fact, laws of mortality (parameter functions) provide a better way determine mortality, being able to give a good fit to empirical mortality curves, mostly because they represent a better means of graduation than discrete mortality representations.

Because we need to focus on mortality representations by differentiable parametric functions, traditional model life tables (tabular representations) of the age pattern is not of interest in this paper.

The results presented in this paper suggest that Gompertz' laws is an appropriate model of mortality to be used for the modern population of Romania.

The paper begins with a presentation of necessary statistic-demographic notions, along with the model that we took into consideration. Within the 2nd section, I focus on the estimation of the parameters, while in the 3rd section we present an application of the model for the population of Romania.

1. Standard statistic-demographic notions

a) Survival probability. Take $P(T_x > t)$ which we will call it the probability of survival for a specific individual of age x after a number of t years, denoted by ${}_t p_x$.

Obviously we say that ${}_0 p_x = 1$.

In general, common laws of mortality use as upper range the age of 100 years, denoted by ώ. Thus we can say: ${}_t p_x = 0$ if $t > ώ - x$.

b) Probability of death. Let ${}_{t|\,t'}q_x$ represent the probability that a specific individual of age x, to die at the age between $x+t$ and $x+t+t'$, meaning:

$$ {}_{t|\,t'}q_x = P(t < T_x \le t + t') = \frac{P(x + t < T_x < x + t + t')}{P(T_x > x)} \quad (1.1) $$

Thus, we can say:

$$ {}_t p_x = {}_{t|\,t'}q_x + {}_{t+t'}p_x \quad (1.2) $$

from where we can establish a connection between the survival probability and the probability of death:

$$ {}_{t|\,t'}q_x = {}_t p_x - {}_{t+t'}p_x \quad (1.3) $$

For easier understanding we take: ${}_t q_x = {}_{0|\,t}q_x$

The next relations are a natural evolution from the above relations:

$$ {}_t q_x = {}_0 p_x - {}_{0+t} p_x = 1 - {}_t p_x \quad (1.4) $$

$$ q_x = {}_1 q_x = 1 - p_x \quad (1.5) $$

$$ {}_{t+t'} p_x = {}_t p_x \cdot {}_{t'} p_{x+t} \quad (1.6) $$

The probability of death q_x is also known under the name of annual rate of mortality or annual coefficient of mortality.

The following are the additional definitions of standard life table functions:

- The entry l_x, i.e. number of survivals (in the life tables), shows the number of survivors of that birth cohort at each succeeding exact integral age.
- The entry d_x, shows the number of deaths that would occur between succeeding exact integral ages among members of the cohort.
- The entry denoted L_x gives the number of person-years lived between consecutive exact integral ages x and x+ 1 and T_x gives the total number of person-years lived beyond each exact integral age x by all members of the cohort.
- The final entry in the life table, $\bar{e}_x$ represents the average number of years of life remaining for members of the cohort still alive at exact integral age x, and is called the life expectancy.

The l_x entry in the life table is also useful for determining the age corresponding to a specified survival rate from birth, which is defined as the age at which the ratio of l_x to 100000 is equal to a specified value between 0 and 1.

The life table functions l_x, d_x, L_x, T_x, and $\bar{e}_x$ are being calculated as follows:

l_0	$= 100000$	
d_x	$= l_x * {}_1q_x$	x = 1, 2, 3,...
l_x	$= l_{x-1} * (1 - {}_1q_{x-1})$	x = 1, 2, 3,...
L_0	$= l_0 - {}_1f_0 * d_0$	
L_x	$= l_x - 1/2 * d_x$	x = 1, 2, 3,...
T_x	$= L_x + L_{x+1} + L_{x+2} + \cdots . + L_{84}$	x = 0, 1, 2, 3,...
$\bar{e}_x$	$= T_x / l_x$	x = 0, 1, 2, 3,...

The model we considered in this paper is an improved version of the Gompertz' law of mortality: $\mu_{b,c}(x) = \frac{c}{b} e^{cx}, \ \ x \geq 0, \ b > 0, \ c \in \mathbb{R}^*$

2. Estimating the parameters from the laws of mortality

2.1 Improved Gompertz' law of mortality

$$\mu_{b,c}(x) = \frac{c}{b} e^{cx}, \ \ x \geq 0, \ b > 0, \ c \in \mathbb{R}^* \quad (2.1)$$

If $c > 0$ we say that Gompertz repartition is of increasing failure rate type (theory of reliability), if $c < 0$ is of decreasing failure rate. From (2.1) we can deduce the corresponding survival function, which looks like:

$$\underline{F}_{b,c}(x) = e^{-\int_0^x \mu(t)dt} \Leftrightarrow \underline{F}_{b,c}(x) = e^{-\frac{e^{cx}-1}{b}},\ x \geq 0 \tag{2.2}$$

along with its density

$$f_{b,c}(x) = -\left(\underline{F}_{b,c}\right)'(x) = \left(\frac{c}{b}\right) e^{cx-\frac{e^{cx}-1}{b}} \tag{2.3}$$

We can drop the constant term $\frac{c}{b}$, by multiplying $f_{b,c}(x)$ with $\frac{b}{c}$. By doing this, we just simplify future computations with the density function, and then dropping it altogether as it is not of interest.

$$f_{b,c}(x) = -\frac{b}{c}\left(\underline{F}_{b,c}\right)'(x) = e^{cx-\frac{e^{cx}-1}{b}} \tag{2.4}$$

The derivative of the density function will look like:

$$\left(f_{b,c}\right)'(x) = \left(c - \frac{c}{b}e^{cx}\right)e^{cx-\frac{e^{cx}-1}{b}} \tag{2.5}$$

In order to establish the solution of this function, we must take into consideration the initial conditions of the model, namely $x \geq 0,\ b > 0,\ c \in \mathbb{R}$.

We can drop the exponential function after the parenthesis because it is obviously positive, as it will not influence the final sign of our function nor will it help providing a solution. After that we extract the common term c, to look like:

$$\left(f_{b,c}\right)'(x) = c\left(1 - \frac{e^{cx}}{b}\right) \tag{2.6}$$

If we can find a solution for this function, we will fix it as the mode for our survival data. Thus we reach the following equation:

$$1 - \frac{e^{cx}}{b} = 0 \ \Leftrightarrow\ e^{cx} = b \tag{2.7}$$

this is so far in line with the initial condition as the left term of the (2.7) is an exponential function and the right term is b, which we already know it to be positive. Further we solve this with the help of natural logarithm:

$$cx = \ln b,\ b > 0 \ \Leftrightarrow x = \frac{\ln b}{c} \tag{2.8}$$

This leads to the conclusion that the final value of the mode is:

$$Mode\left(\Gamma_{b,c}\right) = \frac{\ln b}{c} \tag{2.9}$$

The reason for determining Mode, and further on the Median, is because we lack a methodology to compute the expected value of the Gompertz repartition.

Instead, we can compute its quantile, which will benefit us in the process of determining b and c parameters.

Let m_α represent a solution for the next equation:

$$F_{b,c}(x) = e^{-\alpha} \quad (2.10)$$

Thus

$$m_{ln2} = Median(\Gamma_{b,c}) \quad (2.11)$$

If we apply natural logarithm in the (2.10) equation, we can find the value of m_α:

$$-\frac{e^{cx}-1}{b} = -\alpha \Leftrightarrow e^{cx} = 1 + b\alpha \Leftrightarrow m_\alpha = \frac{ln(1+b\alpha)}{c} \quad (2.12)$$

$$Median(\Gamma_{b,c}) = \frac{ln(1+b\alpha)}{c} \quad (2.13)$$

What is of interest here is that the relation between Median and Mode, depends solely on one parameter, that is b:

$$Median(\Gamma_{b,c})/Mode(\Gamma_{b,c}) = \frac{ln(1+b\alpha)}{lnb} \quad (2.14)$$

If we want, we can generalize the method for determining the parameters, as I present it next. We start by taking 2 values for x, that are significant to us, let's say x_1 and x_2:

$$\begin{cases} F_{b,c}(x_1) = F_1 \\ F_{b,c}(x_2) = F_2 \end{cases} \quad (2.15)$$

where b and c represent the solutions. We replace with $F_1 = e^{-lnm}$ $F_2 = e^{-lnm'}$ and rewrite the (2.15) system to get the following:

$$\begin{cases} e^{-\frac{e^{cx_1-1}}{b}} = e^{-lnm} \\ e^{-\frac{e^{cx_2-1}}{b}} = e^{-lnm'} \end{cases} \Leftrightarrow \begin{cases} -\frac{e^{cx_1}-1}{b} = -lnm \\ -\frac{e^{cx_2}-1}{b} = -lnm' \end{cases} \quad (2.16)$$

Choosing lnm and lnm' was preferable to m and m' for easier computation over the generalized method. Moving forward, if we try to eliminate c from each equation of the (2.16) system, we get the following:

$$\begin{cases} e^{cx_1} = blnm + 1 \\ e^{cx_2} = blnm' + 1 \end{cases} \Leftrightarrow \begin{cases} cx_1 = \ln(1 + blnm) \\ cx_2 = \ln(1 + blnm') \end{cases} \quad (2.17)$$

Now, if we divide the equations from the final form of (2.17) system, we get the following relationship:

$$\frac{x_1}{x_2} = \frac{\ln(1+blnm)}{\ln(1+blnm')} \quad (2.18)$$

where, values of $x_j,\ j = 1,2$ as well as values of m and m' are known; thus the parameter to be determined remains b.

Let's say that $x_1 < x_2$; one can choose x_1 to correspond to the empirical median, which is about age 78, and $x_2 = 100$ (or any last recorded age, in the series, available).

The value of 78 is taken from the life tables, for Romania in the year 2012. (Source: Eurostat database, life tables: national data, demo_mlifetable)

As we want to set a generalized method, one can look in the life tables and extract this value at the age where we have half of the cohort of 100000. Also remember that $m = ln2$ and $m' = -ln\underline{F}(100)$.

3. Application on Romania using Gompertz' law of mortality

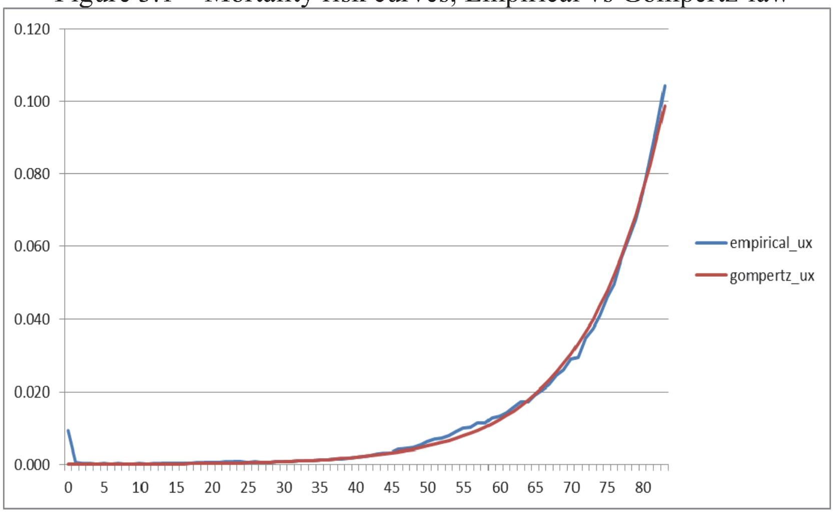
Continuing the work from 2nd section of this paper, we can test the methods using data for Romania, extracted from life tables, in 2012 (Annex 1)

Gompertz's law: In (2.15) equation, we replace m with $ln2$ and m' with the specific value for l_{84} extracted from the life tables (see Annex 1). Taking into consideration the method described in section 2, and replacing $x_1 = 78$, $x_2 = 84$, $l_{84} = 30{,}314$ in (2.18) and (2.16) we determine the parameters: $b = 1678.86$, respectively $c = 0.0905$

After determining the values of both parameters for both laws, we can estimate the survival function and compare it with the empirical one. The most important aspect, here, is to determine the distance between them. The smaller this distance is the better for our model.

In figure 3.1 we can see the representation of both curves, and the fact that Gompertz' curve is very well fitted to the empirical one. We set empirical_ux to represent the empirical data extracted form life tables and gompertz_ux as fitted curve according to Gompertz' law of mortality.

Figure 3.1 – Mortality risk curves, Empirical vs Gompertz'law



Source: Eurostat, demo_mlifetable (author's representation)

4. Conclusions

According to the data determined in section 3 of this paper (see figure 3.1), we can say that Gompertz's has found a way to evaluate the trend of mortality, for a specific population. With this in mind, we can continue and compute the risk of mortality $\mu(x)$. As we already know, Gompertz'law of mortality doesn't fully model certain aspects of mortality trends (infancy and advanced ages risks of mortality), neither did he take into account that not all people die of old age. Putting that aside, the model is still surprisingly well fitted to data available from the last 100 years. Section 2 provided us with the means to calculate certain statistic-demographic indicators, specified usually in life tables, but without relying simply on empirical data. Moreover, due to the lack of data for more advanced ages (over 84 years, and even scarce for over 100 years), we can determine our on data based on the models presented in this study. The same methodology can be also applied in the case of Makeham's curve.

We can confirm the initial hypothesis of Gompertz that the risk of mortality is dependent only on age and location (data in life tables are specific to a country or region). With this we can search for specific patterns or disparities between regions (Diaconescu, 2013, 2014), highlighting possible mortality dominance between population of different sexes, regions or both. We can than start to make a demographic profile related to the probability of death correlated with the average life expectancy and other economic indicators.

Selective references

1. Beard, R.E. (1971), Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes, in W. Brass (ed.), Biological Aspects of Demography, Taylor & Francis Ltd., London
2. Cenuşă, G.; Burlacu, V. (2000) Bazele matematice ale asigurarilor, Ed. Teora, Bucuresti;
3. Diaconescu T. (2014) Mortality Dominance at regional level. Study Case: Romania (to be published)
4. Gnedenko, B., Beliaev Y., Soloviev A (1972) Methodes Mathematiques dans la Theorie de la Fiabilite, Mir;
5. 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", in Philosophical Transactions of the Royal Society of Vol. 36, pp. 513–585, W. Nicol, London.
6. Gompertz, B. (1872) "On one uniform law of mortality from birth to extreme old age, and on the law of sickness", Journal of the Institute of Actuaries, vol. 16, pp. 329–344.

7. Jeune, B. and J.W. Vaupel (1995), Exceptional Longevity: From Prehistory to the Present. Odense Monographs on Population Aging, 2. Odense University Press, Odense, Denmark
8. Kannisto, V., J. Lauritsen, A.R. Thatcher, and J.W. Vaupel (1994), Reductions in mortality at advanced ages: several decades of evidence from 27 countries, Population and Development Review 20:793-870.
9. Kannisto, V.(1996), The Advancing Frontier of Survival: Life Tables for Old Age.Odense Monographs on Population Aging, 3. Odense University Press, Odense, Denmark.
10. Makeham, W.M. (1867) "On the law of mortality", Journal of the Institute of Actuaries, vol. 13, pp. 325–358.
11. http://epp.eurostat.ec.europa.eu/
12.*Selective ideas from doctoral thesis: "Economic Impact of Ageing Population in EU", coordinator Acad. Lucian Liviu Albu., Institute for Economic Forecasting, National Institute for Economic Research "Costin C. Kiritescu" , Romania Academy, Bucharest.

ANNEX 1

Age	Survivors	empirical_ux	gompertz_ux	Age	Survivors	empirical_ux	gompertz_ux
0	100000	0.00920	0.00005	42	96092	0.00254	0.00241
1	99080	0.00070	0.00006	43	95848	0.00289	0.00264
2	99011	0.00041	0.00006	44	95571	0.00315	0.00289
3	98970	0.00026	0.00007	45	95270	0.00308	0.00316
4	98944	0.00018	0.00008	46	94977	0.00422	0.00346
5	98926	0.00028	0.00008	47	94576	0.00426	0.00379
6	98898	0.00018	0.00009	48	94173	0.00464	0.00415
7	98880	0.00024	0.00010	49	93736	0.00537	0.00454
8	98856	0.00020	0.00011	50	93233	0.00611	0.00498
9	98836	0.00017	0.00012	51	92663	0.00690	0.00545
10	98819	0.00028	0.00013	52	92024	0.00719	0.00596
11	98791	0.00021	0.00015	53	91362	0.00798	0.00653
12	98770	0.00027	0.00016	54	90633	0.00900	0.00715
13	98743	0.00027	0.00017	55	89817	0.00996	0.00782
14	98716	0.00041	0.00019	56	88922	0.01026	0.00856
15	98676	0.00033	0.00021	57	88010	0.01135	0.00937
16	98643	0.00045	0.00023	58	87011	0.01145	0.01026
17	98599	0.00047	0.00025	59	86015	0.01278	0.01123
18	98553	0.00062	0.00027	60	84916	0.01324	0.01230
19	98492	0.00062	0.00030	61	83792	0.01425	0.01346
20	98431	0.00068	0.00033	62	82598	0.01593	0.01474
21	98364	0.00057	0.00036	63	81282	0.01733	0.01613
22	98308	0.00079	0.00039	64	79873	0.01714	0.01766
23	98230	0.00086	0.00043	65	78504	0.01903	0.01934
24	98146	0.00088	0.00047	66	77010	0.02046	0.02117
25	98060	0.00068	0.00052	67	75434	0.02193	0.02317
26	97993	0.00082	0.00057	68	73780	0.02460	0.02537
27	97913	0.00068	0.00062	69	71965	0.02594	0.02777
28	97846	0.00069	0.00068	70	70098	0.02899	0.03040
29	97778	0.00077	0.00074	71	68066	0.02937	0.03328
30	97703	0.00088	0.00081	72	66067	0.03474	0.03643
31	97617	0.00085	0.00089	73	63772	0.03754	0.03988
32	97534	0.00097	0.00098	74	61378	0.04098	0.04366

Age	Survivors	empirical_ux	gompertz_ux	Age	Survivors	empirical_ux	gompertz_ux
33	97439	0.00106	0.00107	75	58863	0.04604	0.04780
34	97336	0.00111	0.00117	76	56153	0.04958	0.05233
35	97228	0.00131	0.00128	77	53369	0.05681	0.05728
36	97101	0.00133	0.00140	78	50337	0.06178	0.06271
37	96972	0.00153	0.00153	79	47227	0.06752	0.06865
38	96824	0.00163	0.00168	80	44038	0.07444	0.07515
39	96666	0.00178	0.00184	81	40760	0.08425	0.08227
40	96494	0.00201	0.00201	82	37326	0.09323	0.09006
41	96300	0.00216	0.00220	83	33846	0.10436	0.09859
				84	30314		