

Expanding an Abridged Life Table Using the Heligman-Pollard Model

Rose Irnawaty Ibrahim

Institute of Mathematical Sciences, University of Malaya
50603 Lembah Pantai, Kuala Lumpur, Malaysia
e-mail: rose.irnawaty@um.edu.my

Abstract An estimation procedure for calculating the one-year probabilities of dying from five year ones given in an abridged life table will be proposed in the study. The expansion technique used in the study is the Heligman-Pollard model. The evaluation of this technique and several other techniques for expanding an abridged life table will be explained in this paper. Since the model involves nonlinear equations that are explicitly difficult to solve, the Matrix Laboratory Version 7.0 (MATLAB 7.0) software will be used in the study. A nonlinear least squares algorithm with the capability of approximating numerically all derivatives was used in order to estimate the parameters of the model. This algorithm is based upon a modification of the Gauss Newton iteration procedure, known as Levenberg-Marquardt iteration procedure. The empirical data sets of Malaysia population for the period of 1991-2000 and for both genders will be considered.

Keywords Abridged life table, Heligman-Pollard model, Modification of the Gauss Newton iteration, Levenberg-Marquardt iteration.

1 Introduction

A life table is a statistical device used by actuaries, demographers, public health employees and many others to present the mortality experience of a population. It is also referred to as the mortality table. The abridged life tables are more often used, in which each age is presented in groups, which are 0-1, 1-5, 5-10, 10-15 and so on. The use of abridged life tables has expanded because mortality data are usually available and sufficiently accurate in the form of rates for 5-year age groups and not for each individual age. The main reason for providing data in an abridged form is related to the phenomenon of age heaping caused by age misstatements in data registration. Another reason is of the incomplete and unstable documentation of vital statistics and therefore the quality of the data may not permit computation of a complete life table. Thus, it often happens that only an abridged life table is available when a complete life table is needed. Hence it is important to construct a complete life table given the abridged one.

Several methods have been suggested in the literature to construct a complete life table given the abridged life table. The methods that have been extensively used are interpolation formula as in King [9] and Beers [1]. However, an interpolation method that is still being

used is the six-point Lagrangian interpolation formula by Elandt-Johnson [5] which is applied to the $l(x)$ values in an abridged life table. This formula expresses each non-tabulated value of $l(x)$ as a linear combination of six particular polynomials in x , each of degree five. The formula is given as

$$l(x) = \sum_{i=1}^6 \frac{\prod_{j \neq i} (x - x_j)}{\prod_{j \neq i} (x_i - x_j)} l(x_i)$$

where x_1, x_2, x_3, x_4, x_5 and x_6 are the tabular ages nearest to x . This method provides good approximations for adult mortality but it is less accurate for the early childhood ages. However the application of Lagrangian formula is limited to the ages less than 75 years while literature proposes other complementary methods for the ages greater than 75 years old.

Valaoras [15] was developed an alternative procedure for breaking abridged life tables into one year age groups. He used this procedure to produce the official complete life table for the Greek male and female populations. This method is based on the five years mortality rates expressed as

$${}_5m_x = \frac{{}_5d_x}{{}_5M_x}$$

where ${}_5d_x$ is the number of deaths in the age interval $[x, x + 5)$ and ${}_5M_x$ is the mean population in the age interval $[x, x + 5)$.

Then, the one year probabilities of dying $q_7, q_{12}, q_{17}, \dots$ are calculated by the use of the approximate formula as

$$q_{x+2} = \frac{2 \times {}_5m_x}{2 + {}_5m_x}$$

Thus, to estimate the complete set of q_x values for $x \geq 5$, the formula that will be used is given by

$$\frac{q_x}{K^x} = a + bx + cx^2 + dx^3$$

By using the values of $q_7, q_{12}, q_{17}, \dots$, the parameters K, a, b, c and d will be estimated using least squares technique. For detail refer to Valaoras [15]. However, this method is not adequate for approximating early childhood mortality, which is for $x < 5$ and for earlier adult ages, where $18 < x < 40$.

A recent attempt to represent mortality over the course of the entire life span using a single analytical expression has been made in Australia by Heligman and Pollard [7]. Several applications of Heligman and Pollard model on a wide variety of mortality experiences have been used in United State of America by Mode and Busby [14] and in Sweden by Hartmann [6]. Hartmann [6] concluded that Heligman and Pollard model is the best existing demographic model of mortality at all ages and is an efficient means of generating life tables model, for example for use of population projection. The Heligman and Pollard model also was discussed by Kostaki [10]. He concluded that this model provides quite a satisfactory representation of the age pattern of mortality. He also concluded that this model provides a new way to expand a life table through direct estimation of the complete set of probabilities of dying, q_x , having the abridged ${}_nq_x$ values as the starting point.

2 Method of Expanding an Abridged Life Table to a Complete Life Table

As mentioned above, the best existing model of mortality at all ages and an efficient mean of generating life table model is the model proposed by Heligman and Pollard [7]. The mathematical function of Heligman and Pollard model is given by

$$\frac{q_x}{p_x} = A^{(x+B)^C} + D * \exp \left(-E \left(\ln \left(\frac{x}{F} \right) \right)^2 \right) + GH^x \quad (1)$$

where q_x is the probability that an individual who has reached age x will die before reaching age $x + 1$ while $p_x = 1 - q_x$ and A, B, C, D, E, F, G, H are the positive parameters to be estimated.

The model contains three terms, each representing a distinct component of mortality. The first term is a rapidly decreasing exponential function and reflects the fall in mortality at the infant and early childhood ages, which are at ages less than 10 years. This component of mortality has three parameters: A , which is nearly equal to q_1 , measures the level of mortality. C measures the rates of mortality decline in childhood (the higher the value of C , the faster the mortality decreases with an increasing age), while B is an age displacement to account for infant mortality (when $B = 0$, $q_0 = 0.5$ for any values of A and C).

The second term is a function similar to the lognormal density and reflects the middle life mortality. It reflects the accident mortality for males and accident plus maternal mortality for the females, often referred to in the demographic literature as the accident “hump”. The accident term has three parameters: F indicating the location, E representing the spread and D the severity. And finally, the last term is a Gompertz exponential function which reflects the rise in mortality at the adult ages, which is ages greater than 40 years old.

In order to expand an abridged life table to a complete life table, the parameters of the Heligman and Pollard model need to be estimated. Letting $Q(x, c)$ to denote the right hand side of (1), it is reasonable to expect that the parameters can be estimated by minimizing the sum of squares

$$\sum_x \left(\frac{\hat{q}_x}{\hat{p}_x} - Q(x, c) \right)^2$$

with respect to $c = (A, B, C, D, E, F, G, H)$. Hartmann [6] explained that this procedure cannot be used because it will give a negative estimate of B , which is not permitted. Therefore, it is important to follow the procedure given by Heligman and Pollard [7] in order to estimate these parameters. Since the purpose of the study is to analyse the mortality rates for person aged 10 years and above, the first term of Heligman and Pollard model is negligible. Therefore, (1) can be simplified and becomes

$$\frac{q_x}{p_x} = D * \exp \left(-E \left(\ln \left(\frac{x}{F} \right) \right)^2 \right) + GH^x. \quad (2)$$

From Equation (2), it is clear that we only need to estimate five parameters of Heligman and Pollard model, which are D, E, F, G and H .

In order to estimate these parameters, we start by collecting the abridged ${}_nq_x$ values which were given in the abridged life tables for the Malaysian population. In the study, the abridged life tables for the Malaysian population in year 1991 until year 2000 for both

males and females were be collected. Then these values will be fitted to (2). The fit is judged on the basis of the sum of the squared differences or ‘residuals’ between the input data points and the function values, evaluated at the same places. Thus, fit is used to find a set of parameters that ‘best’ fits the data to user-defined function. A parameter is a user-defined variable that fit will adjust, which is an unknown quantity in the function declaration. However, the sum of squares to be minimized in the study was given by

$$S^2 = \sum_x \left(\frac{n\hat{q}_x}{nq_x} - 1 \right)^2 \quad (3)$$

Often the function to be fitted will be based on a model or theory that attempts to describe or predict the behaviour of the data. Then ‘fit’ can be used to find values for the free parameters of the model, to determine how well the data fits the model, and to estimate a standard error for each parameter. In the study, the function to be fitted will be based on a Heligman and Pollard model.

Let us use $F(x; c)$ to denote the right hand side of (2) and our model will become

$$\frac{q_x}{p_x} = F(x; c).$$

Then, for the one year odds of dying, we obtain

$$q_x = \frac{F(x; c)}{1 + F(x; c)}.$$

From the relation of

$${}_nq_x = 1 - \prod_{i=0}^{n-1} (1 - q_{x+i}),$$

the model for the death probabilities in the abridged life table becomes

$${}_n\hat{q}_x = 1 - \prod_{i=0}^{n-1} \left(1 - \frac{F(x+i; c)}{1 + F(x+i; c)} \right).$$

Since this minimization problem involves nonlinear equations that are difficult to solve explicitly, the Matrix Laboratory Version 7.0 (MATLAB 7.0) software will be used in the study. A nonlinear least squares algorithm with the capability of approximating numerically all derivatives was used in order to estimate the parameters of (2). The nonlinear least squares problem is to find a vector of parameter values to minimize the sum of squares, $S^2(3)$. However, in the study the Levenberg-Marquardt iteration procedure will be used to get the estimated values of these parameters (Levenberg [11] and Marquardt [13]) by using a damped Gauss-Newton iteration procedure. The Levenberg-Marquardt algorithm provides a numerical solution to the mathematical problem of minimizing a sum of squares of several, generally nonlinear functions that depend on a common set of parameters. Like other numeric minimization algorithms, the Levenberg-Marquardt algorithm is an iterative procedure. The Levenberg-Marquardt algorithm interpolates between the Gauss-Newton algorithm and the method of gradient descent. The Levenberg-Marquardt algorithm is more robust than the Gauss-Newton algorithm since in many cases it finds a solution even if it starts very far off the global minimum.

The Levenberg-Marquardt algorithm selects the parameter values for the next iteration. The process continues until a preset criterion is met, either (i) the fit has “converged” or (ii) it reaches a preset iteration count limit. The algorithm of the method is presented below, where f is the minimized sum of squares function, J is the Jacobian matrix and τ is chosen by user. This algorithm is not very sensitive to the choice of τ , but as a rule of thumb, one should use a small value, for example $\tau = 10^{-6}$.

```

begin
iter := 0; v := 5; x := xo; A := JTJ; g := JTf
found := ( $\|g\|_{\infty} \leq \varepsilon_1$ );  $\lambda := \tau * \max \{a_{ii}\}$ 
while(not found) and (iter < itermax)
iter := iter + 1; Solve (A +  $\lambda I$ ) hlm = -g
if  $\|h_{lm}\| \leq \varepsilon_2 (\|x\| + \varepsilon_2)$ 
found := true
else
xnew := x + hlm;  $\delta := (F(x) - F(xnew)) / (L(0) - L(h_{lm}))$ 
if  $\delta > 0$ 
x := xnew; A := JTJ; g = JTf
found := ( $\|g\|_{\infty} \leq \varepsilon_1$ )
 $\lambda := \lambda * \max \{1/3, 1 - (2\delta - 1)^3\}$ ; v := 5
else
 $\lambda := \lambda * v$ ; v := 5 * v;
end

```

Then by referring to the algorithm above, the program for this algorithm will be formulated by using the MATLAB 7.0 software. Result of estimated parameters in Heligman-Pollard model in year 1991 until year 2000 for both males and females will be discussed in the next section. By inserting these estimated values of the parameters into (2), the one year probabilities of dying is represented by the mortality rates at age x , that is q_x can be calculated.

3 Results and Discussion

Tables 1 and 2 give results of estimated parameters in Heligman-Pollard model for period 1991 to 2000 for males and females, respectively.

From tables above, we found that estimated parameters D , E , and F are around 0.001, 9 and 21 respectively for males and estimated parameters D , E , and F are around 0.0004, 3 and 19 respectively for females. And we also found that estimated parameter G for males is around 0.0001 and females have consistently declined. Also we see that the values of G were higher for males than for females over a period of study. The Gompertz parameter H is almost constant (about 1.1) over the period of study.

Then by inserting these parameters into (2), the mortality rates at age x can be calculated. Based on our results, we drew graphs for mortality rates at age $x(q_x)$ against age (x) for both males and females in year 1991 until year 2000 using Microsoft Excel 97. Both graphs are displayed and discussed below. At present it is a common practice to plot graphs of $\log q_x$ against x , but at higher ages these give a curve, which is not very informative. However, when a simple Heligman and Pollard law holds, a graph of $\log (q_x/p_x)$ for ages

Table 1: Result of Estimated Parameters in Heligman-Pollard Model for Period 1991 to 2000 for Males

Parameters	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>
1991	0.001267	8.98175	20.95931	0.00008338	1.094737
1992	0.001302	8.98223	20.96418	0.00008005	1.095292
1993	0.001347	8.99988	20.99980	0.00007761	1.095811
1994	0.001415	8.98157	20.96397	0.00007845	1.095689
1995	0.001468	8.97995	20.95867	0.00008459	1.094303
1996	0.001430	8.97882	20.96778	0.00009450	1.092386
1997	0.001397	8.99985	20.99982	0.00009531	1.091640
1998	0.001311	8.99615	20.99559	0.00010131	1.090534
1999	0.001275	8.99983	20.99980	0.00010020	1.090360
2000	0.001502	8.99997	20.99999	0.00008889	1.091765

Table 2: Result of Estimated Parameters in Heligman-Pollard Model for Period 1991 to 2000 for Females

Parameters	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>
1991	0.0004087	2.99995	18.99998	0.00002906	1.107088
1992	0.0003989	2.99996	18.99998	0.00002767	1.107684
1993	0.0003917	2.99995	18.99997	0.00002558	1.108766
1994	0.0003970	2.99995	18.99997	0.00002438	1.109367
1995	0.0003880	2.99996	18.99997	0.00002310	1.110110
1996	0.0003853	2.99996	18.99997	0.00002333	1.109770
1997	0.0003750	2.99996	18.99996	0.00002192	1.110510
1998	0.0003675	2.99993	18.99995	0.00002170	1.110425
1999	0.0003656	2.99992	18.99994	0.00002114	1.110467
2000	0.0003759	2.99992	18.99995	0.00001963	1.111655

40 and above will be a straight line. Therefore, in order to check if our results follow the Heligman and Pollard law, we drew graphs of $\log (q_x/p_x)$ against x for both males and females aged 40 and above in year 1991 to 2000 using Microsoft Excel 97. Both graphs are presented and discussed below.

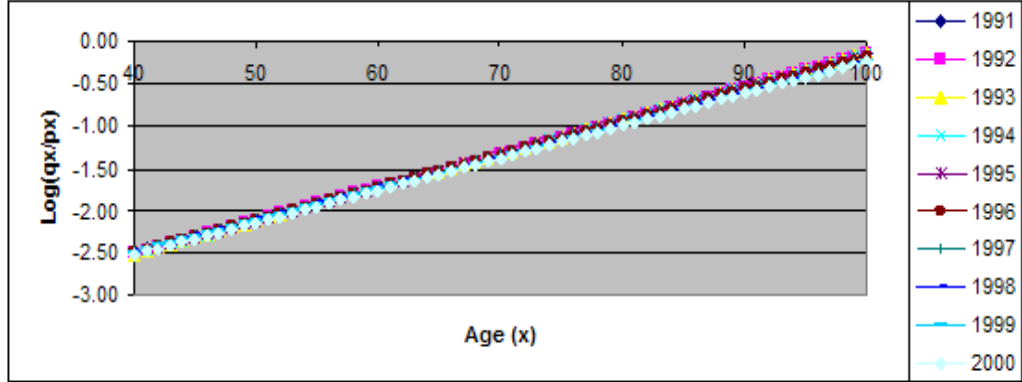


Figure 1: Graph $\text{Log} (q_x/p_x)$ Against Age (x) for Males Aged 40 and Above in Year 1991 to 2000

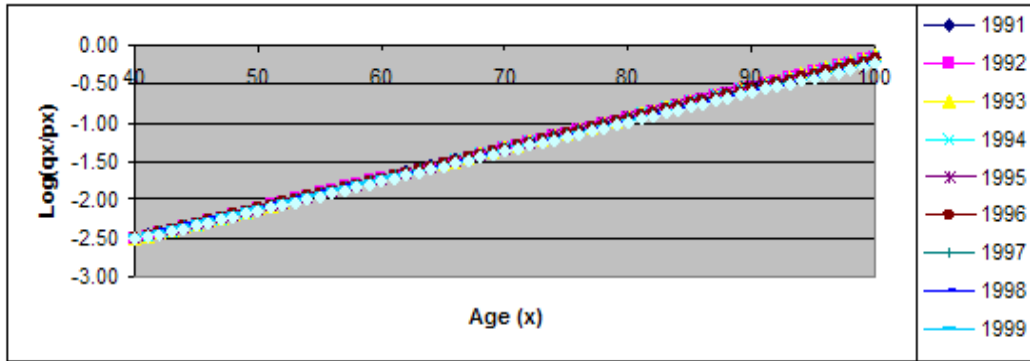


Figure 2: Graph $\text{Log} (q_x/p_x)$ Against Age (x) for Females Aged 40 and Above in Year 1991 to 2000

From Figure 1 and Figure 2, we can see clearly that both graphs are straight lines. Therefore, we can conclude that our results follow the Heligman and Pollard law.

Generally, mortality is not something that changes very drastically over a short period of time and thus one set of mortality rates is sufficient for the analysis and computations in the study. In order to get the one set of mortality rates, we fit the model to the data, which is q_x in 1991 to 2000 using nonlinear regression to determine the best estimated values of parameters D, E, F, G and H use SPLUS software. By using this software, we obtained the equations for fitted nonlinear regression equation for males and females, respectively as

follows:

$$\frac{q_x}{p_x} = 0.00137235 \times \exp(-8.94483 \times (\ln(x/20.9846))^2) + (0.0000879068 \times (1.09331^x))$$

and

$$\frac{q_x}{p_x} = 0.000385398 \times \exp(-2.98580 \times (\ln(x/19.0065))^2) + (0.0000236727 \times (1.10954^x))$$

Based on the results of the mortality rates at age x for both males and females obtained above, we drew graphs of mortality rate against age for both males and females using Microsoft Excel 97. To make comparison, we drew graphs of mortality rate against age for both genders at difference ages. The graph is presented in Figures 3, 4 and 5, and discussed below.

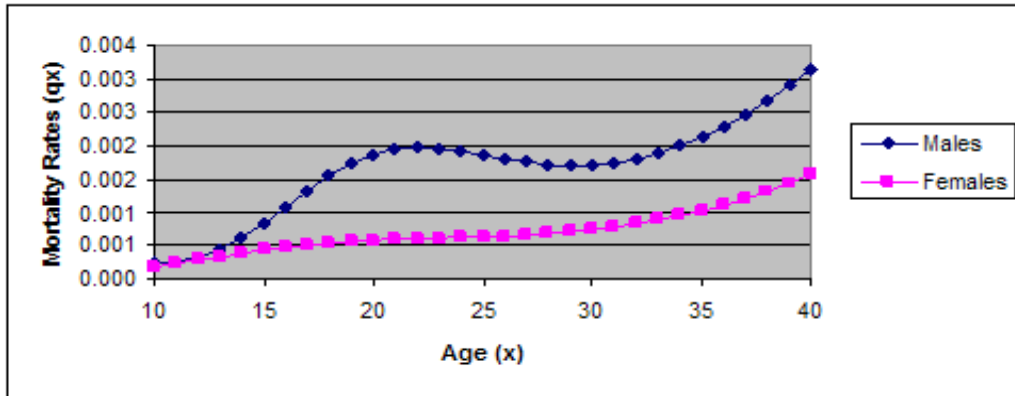


Figure 3: Graph Mortality Rate Against Age for Males and Females at Aged 10 to 40

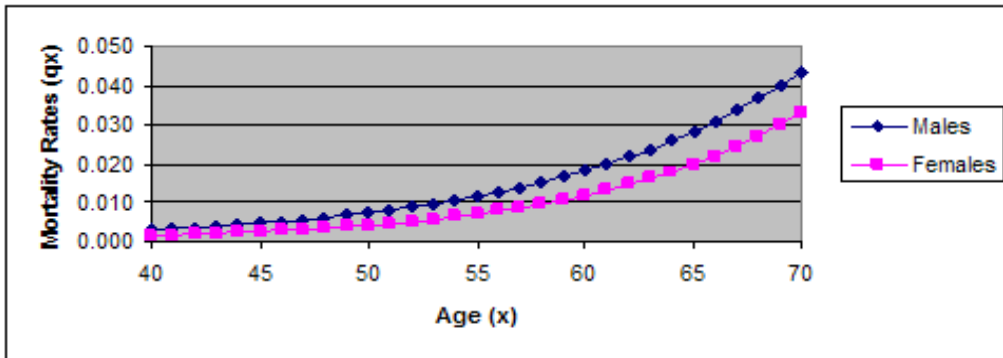


Figure 4: Graph Mortality Rate Against Age for Males and Females at Aged 40 to 70

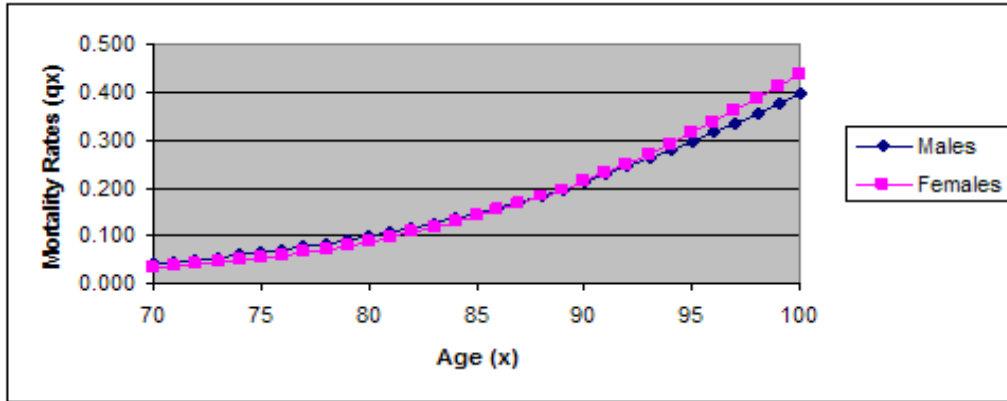


Figure 5: Graph Mortality Rate Against Age for Males and Females at Aged 70 to 100

As can be seen from these figures, we found that the graphs for Females exhibit lighter mortality at ages 10 to 87 compared to that for Males. We also can see clearly that the mortality rates increase with age and it is widely accepted. Additionally, the rate of mortality increases almost exponentially with age for adult as mentioned by Horiuchi and Wilmoth (1998) and Brown (1997, 1988). However, from the study, (refer Figure 4 and Figure 5) we found that starting from age of 40 the mortality rates increase almost exponentially with age. This is because this mortality rates follows the last term of the Heligman-Pollard model, which is a Gompertz exponential function and reflects the rise in mortality at the adult ages.

Most mortality studies clearly indicate that female mortality is significantly lower than male mortality, and the difference may be increasing. From our analysis we see that the mortality rates for females are different than that for males, and the difference is generally increasing until it reaches its peak at the oldest age category. Basically, females have experienced smaller mortality, even though their morbidity rate or sickness rate is higher than that of males. Beside that, there is an accepted difference in life expectancy at birth between the two sexes that favours females. In other words, the life expectancy among population is increasing due to advances in medical knowledge and facilities. In addition, when we compared the actual data (data on ${}_nq_x$) with the estimated ${}_nq_x$, we found that there are no significant differences between them.

4 Conclusion

The objective of the study is to analyse the mortality rates for persons aged 10 years and above for both males and females. In order to get the result, we need to expand the ${}_nq_x$ in an abridged life table to q_x . Several methods have been suggested in the literature to construct a complete life table given the abridged life table. Since found that there are no significant differences between the actual data (data on ${}_nq_x$) and the estimated ${}_nq_x$, we can conclude that the Heligman and Pollard model is the best existing demographic model of mortality at all ages and is an efficient means of generating life tables model, for example

for use of population projection. Also results have shown that the model provides quite a satisfactory representation of the age pattern of mortality.

References

- [1] S.H. Beers, *Six-term Formulae for Routine Actuarial Interpolation*, *R.A.I.A.*, 33(1944), 245–260.
- [2] B. Benjamin & J.H. Pollard, *The Analysis of Mortality and other Actuarial Statistics*, second edition, William Heinemann Ltd, London, 1980.
- [3] R.L. Brown, *Theories of Mortality*, Education and Examination Committee of the Society of Actuaries, Course 161, Study Note 161–202. Schaumburg, IL: Society of Actuaries, 1988.
- [4] R.L. Brown, *Issues in the Modelling of Mortality at Advanced Ages*, Research Report 97-05, Institute of Insurance and Pension Research, University of Waterloo, Ontario, Canada, 1997.
- [5] R. Elandt-Johnson & N. Johnson, *Survival Models and Data Analysis*, John Wiley, New York, 1980.
- [6] M. Hartmann, *Past and Recent Attempts to Model Mortality at All Ages*, *Journal of Official Statistics*, Statistics Sweden, 3(1987), 19–36.
- [7] L. Heligman & J.H. Pollard, *The Age Pattern of Mortality*, *Journal of the Institute of Actuaries*, 107, Part 1(1980), 49–80.
- [8] S. Horiuchi & J.R. Wilmoth, *Deceleration in the Age Pattern of Mortality at Older Ages*, *Demography*, 35(1998), 391–412.
- [9] G. King, *On a Short Method of Constructing an Abridged Life Table*, *Journal of the Institute of Actuaries* 48(1914), 294–303.
- [10] A. Kostaki, *The Heligman-Pollard Formula as a Tool for Expanding an Abridged Life Table*, *Journal of Official Statistics*, 7(1991), 311–323.
- [11] K. Levenberg, *A Method for the Solution of Certain Problems in Least Squares*, *Quart. Appl. Math.* 2(1944), 164–168.
- [12] D. London, *Survival Models and Their Estimation*, Actex Publications, U.S., 1987.
- [13] D. Marquardt, *An Algorithm for Least Squares Estimation on Nonlinear Parameters*, *Journal of Numerical Analysis*, SIAM, 11(1963), 431–441.
- [14] C. Mode & R. Busby, *An Eight Parameter Model of Human Mortality-the Single Decrement Case*, *Bulletin of Mathematical Biology*, 44(1982), 647–659.
- [15] V. Valaoras, *The 1980 Life Tables for Greece*, Publications of the Academy of Athens, 59 (1984), 405–436.