1997

Comments on Some Parametric Models for Mortality Tables

Kam C. Yuen
University of Hong Kong, Hong Kong, kcyuen@hkucc.hku.hk

Comments on Some Parametric Models for Mortality Tables

Kam C. Yuen*

Abstract†

Parametric models for the entire age pattern of mortality have been suggested by Heligman and Pollard (1980) and Carriere (1992). The former is designed to fit the classical mortality pattern while the latter is supported by a statistical theory. Insights into their papers motivate us to consider a variation of the Heligman-Pollard model. We also apply these models to the 1993 Hong Kong Assured Lives Mortality Tables as well as the 1991 Hong Kong Female Life Table. This paper is not intended to construct a better parametric model for mortality tables; the main purpose is simply to provide insights into the potential of these models.

Key words and phrases: *Inverse-Weibull, Inverse-Gompertz, Gompertz, Weibull, mortality*

1 Introduction

The study of parametric models for mortality tables, sometimes referred to as the *law of mortality*, has been of interest to actuaries for many years. A good model can give us a better understanding of the underlying mechanism governing the mortality pattern. Recent development in this topic can be found in Forfar *et al.*, (1988), Renshaw (1991), Tenenbein and Vanderhoof (1980), and Wetterstrand (1981).

*Kam C. Yuen, Ph.D., A.S.A., is an assistant professor in Actuarial Science and Statistics at the University of Hong Kong, Hong Kong. He obtained his Ph.D. in statistics from the University of Calgary, Canada. His papers have appeared in various journals including *Biometrika* and *Journal of Statistical Planning and Inference.*

Dr. Yuen's address is: Department of Statistics, The University of Hong Kong, Pokfulam Road, HONG KONG. Internet address: *kcyuen@hkucc.hku.hk*

†The author is grateful to the referees for many helpful comments. This research was supported by the Committee on Research and Conference Grants of the University of Hong Kong.

A parametric model for mortality tables has many advantages: (i) it involves parameters having demographic and statistical interpretations; (ii) it is applicable to non-integral ages; (iii) it allows comparison among mortality tables by comparing only a few parameters, and; (iv) it provides a ready extrapolation beyond the range of the available data.

The purpose of this paper is to study two parametric models for modeling the pattern of mortality: one proposed by Heligman and Pollard (1980) and the other by Carriere (1992). A brief description of both models is given in Section 2. Insights into their papers and a variant on the Heligman-Pollard model are presented in Section 3. The Heligman-Pollard model is designed to fit the classical pattern of mortality; in some cases a modified version may perform better. In Section 4, we fit the models to the 1993 Hong Kong Assured Lives Mortality Tables presented by the Actuarial Society of Hong Kong (1993) and to the 1991 Hong Kong Female Life Table published by the Census and Statistics Department of Hong Kong (1992). To conclude this paper, we remark on various aspects of these two models.

Finally, the objective of this paper is not to build a better parametric model for mortality tables. Instead, we are interested in exploring modifications to these models that may be better. It is hard to say which model is the best. It all depends on the pattern of mortality, the theory behind the model, and the interpretation of the parameters. Therefore, it is wise to plot the mortality pattern and to consider various aspects of the visible mortality patterns before making a choice.

2 Heligman-Pollard and Carriere Models

We will present only a brief discussion of these models. For more details about the two models, we refer the reader to the original papers.

2.1 The Heligman-Pollard Model

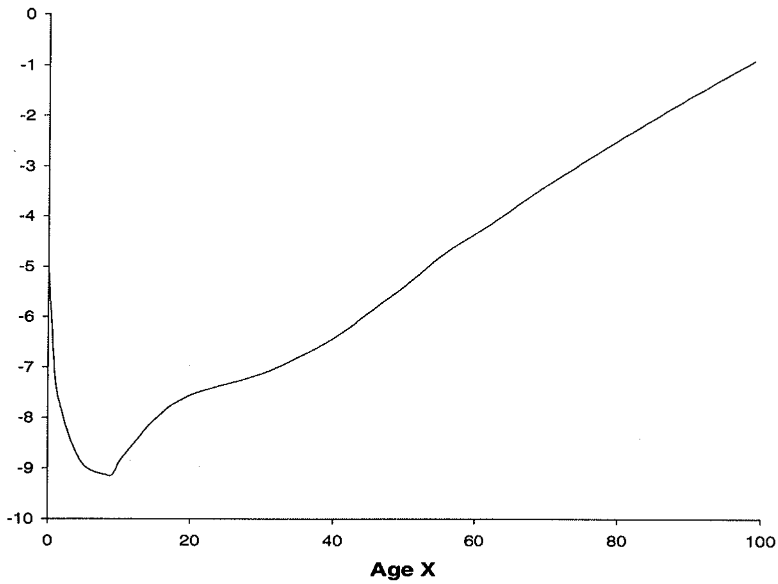
Heligman and Pollard (1980) propose a mathematical expression for the graduation of the pattern of mortality that fits Australian mortality fairly well at all ages. Their law of mortality has the form

$$q_x/(1-q_x) = A^{(x+B)^C} + D\exp\{-E(\ln x - \ln F)^2\} + GH^x \qquad (1)$$

where q_x is the probability that a person age x will die within a year.

Equation (1) contains three terms, each representing a distinct component of mortality. The first term reflects the fall in mortality during childhood. The second term reflects the hump that generally exists between ages 10 and 40. This hump is a consequence of the elevated accident mortality for males and the increased accident mortality plus maternal mortality for females. The third term reflects the exponential pattern of mortality at adult ages.

Figure 1
Hong Kong 1991 Male Table: Plot of $\ln(q_x)$ vs. x



Many mortality tables exhibit the classical pattern suggested by equation (1). Such a pattern is illustrated by plotting $\ln(q_x)$ versus x, using the 1991 Hong Kong Male Life Table; see Figure 1. In Figure 1, we can immediately identify a fall at the early ages, a hump at around 22, and a linear component at the adult ages. The demographic interpretation of the eight parameters in Equation (1) is as follows: A measures the level of mortality in childhood; B is an age displacement to account for infant mortality; C measures the rate of mortality decline in childhood; D, E, and F represent the severity, spread, and location in the accident term, respectively; G represents the base level of mortality at the senior ages; while H reflects the rate of increase of mortality at the adult ages.

2.2 The Carriere Model

Carriere (1992) establishes another parametric model for life tables that also describes the entire age pattern of mortality. This model can be written as a mixture of n survival functions ($S_k(x)$, $k = 1, \ldots, n$), i.e.,

$$S(x) = \sum_{k=1}^{n} \omega_k S_k(x)$$

where the ω's are mixing probabilities with $\sum_{k=1}^{n} \omega_k = 1$. Then, q_x can be evaluated by the relationship

$$q_x = 1 - S(x+1)/S(x).$$

For modeling the classical pattern of mortality, Carriere used $n = 3$, i.e.,

$$S(x) = \omega_1 S_1(x) + \omega_2 S_2(x) + \omega_3 S_3(x) \tag{2}$$

where the parameter ω_1 may be interpreted as the probability that a new life will die during childhood. Similar interpretations apply to ω_2 and ω_3.

Carriere argued that extreme-value survival functions are reasonable models for $S_k(x)$'s. Table 1 summarizes the distributions suggested by Carriere (1992). Note that one can choose either Inverse-Weibull or Inverse-Gompertz to depict the mortality for teenage years, i.e., the accident hump. From Table 1, we note that equation (2) represents an eight-parameter model just like equation (1).

The forms of the distributions in Table 1 look rather different from the standard ones, for example, $\mu_x = GH^x$, for the Gompertz distribution. Carriere claims that this reparametrization provides an insightful statistical interpretation in the sense that $m > 0$ is a measure of location and that $\sigma > 0$ is a measure of dispersion about m. For the Weibull distribution and the Inverse-Weibull distribution, however, m and σ are statistically informative only when σ is small relative to m.

3 Insights and Variations

It is well-known that the third term of equation (1) is the force of mortality of the Gompertz distribution. Helgiman and Pollard also mentioned in their paper that the second term of equation (1) is similar to the lognormal distribution.

Table 1

Component Distributions for Equation (2)

	Childhood ($S_1(x)$)	Teenage ($S_2(x)$)		Adult ($S_3(x)$)
	Weibull	Inverse-Weibull	Inverse-Gompertz	Gompertz
$S(x)$	$\exp\{-(\frac{x}{m})^{\frac{m}{\sigma}}\}$	$1-\exp\{-(\frac{x}{m})^{-\frac{m}{\sigma}}\}$	$\frac{1-\exp\{-e^{-\frac{(x-m)}{\sigma}}\}}{1-\exp\{-e^{\frac{m}{\sigma}}\}}$	$\exp\{e^{-\frac{m}{\sigma}}-e^{\frac{(x-m)}{\sigma}}\}$
$f(x)$	$\frac{\frac{1}{\sigma}(\frac{x}{m})^{\frac{m}{\sigma}-1}}{\exp\{(\frac{x}{m})^{\frac{m}{\sigma}}\}}$	$\frac{\frac{1}{\sigma}(\frac{x}{m})^{-\frac{m}{\sigma}-1}}{\exp\{(\frac{x}{m})^{-\frac{m}{\sigma}}\}}$	$\frac{\frac{1}{\sigma}\exp\{-\frac{x-m}{\sigma}-e^{-\frac{(x-m)}{\sigma}}\}}{1-\exp\{-e^{\frac{m}{\sigma}}\}}$	$\frac{\exp\{\frac{x-m}{\sigma}+e^{-\frac{m}{\sigma}}\}}{\sigma\exp\{e^{\frac{(x-m)}{\sigma}}\}}$
μ_x	$\frac{1}{\sigma}(\frac{x}{m})^{\frac{m}{\sigma}-1}$	$\frac{\frac{1}{\sigma}(\frac{x}{m})^{-\frac{m}{\sigma}-1}}{\exp\{(\frac{x}{m})^{-\frac{m}{\sigma}}\}-1}$	$\frac{\frac{1}{\sigma}\exp\{-\frac{x-m}{\sigma}\}}{\exp\{e^{-\frac{(x-m)}{\sigma}}\}-1}$	$\frac{1}{\sigma}\exp\{\frac{x-m}{\sigma}\}$

Note: Equation (4) uses a different parametrization for the Inverse-Gompertz density.

It appears that Helgiman and Pollard did not recognize the first term, which is equivalent to the three parameter Weibull survival function:

$$S(x) = A^{(x+B)^C} = \exp\left(-(a(x+b))^c\right) \tag{3}$$

for $x > -b$, $a > 0$, and $c > 0$. Therefore, the parameters A, B and C can be rewritten as $\exp(-a^c)$, b, and c, respectively. If we let $b = 0$, then we have the two parameter (standard) Weibull distribution. Thus, the Heligman-Pollard model of equation (1) is related to three distinct lifetime distributions.

From the viewpoint of curve-fitting, the three parameter Weibull is better than the two parameter Weibull because the location parameter b can shift the two parameter Weibull curve back and forth. For example, if equation (1) does not have the parameter B, then q_0 will be fixed at $1/2$ no matter what values A and C may have (the value of G is usually small).

The implicit idea behind the Heligman-Pollard model is that there are three distinct components of human mortality. With this in mind, we may wish to find other functions to replace those in equation (1) provided that they can do the job better. In our opinion, the first and third terms of equation (1) fit extremely well. It is hard to find other functions to supersede them. We may use the Inverse-Gompertz or Inverse-Weibull, however, to handle the second component. For example, the model

$$\frac{q_x}{1-q_x} = A^{(x+B)^C} + D \times \frac{E\ln(\frac{1}{F})\, F^x \exp(-EF^x)}{1-\exp(-E)} + GH^x \tag{4}$$

fits the 1991 Hong Kong Female Life Table better than equation (1); see Tables 4 and 5. For computational and notational convenience, the Inverse-Gompertz density in the second term of equation (4) is reparametrized using E and F instead of the m and σ shown in Table 1. The parameters D, $-\ln(E)/\ln(F), -1/\ln(F)$ can be interpreted as the severity, location, and spread in the accident component. This reparametrization also is used in the numerical examples given in the next section.

Under the Balducci assumption, the function $q_x/(1-q_x)$ is the same as the force of mortality at x. Hence, Carriere construes equation (1) as a total force of decrement that is equal to the sum of three forces of decrement from different causes. This interesting statistical idea still holds for equation (4). Provided that each of the three terms is nonnegative for the range of x and that $\int_0^\infty q_x/(1-q_x)dx = \infty$, other variations of the Heligman-Pollard model share the same interpretation.

Using the parametrization of the Weibull distribution in Table 1, it is possible to gain further insight by focusing on the mean, $m\Gamma(1+\sigma/m)$, and the median, $m(\ln 2)^{\sigma/m}$. Each of these quantities is close to m when σ is small relative to m. With this restriction, the location parameter m and the dispersion parameter σ are statistically informative. The values of σ and m quoted in various applications of equation (2), however, do not conform with the assumption that $\sigma < m$. The same comments apply to the Inverse-Weilbull distribution.

In the theory of lifetime distribution (Lawless 1982), the parameter c in equation (3) or m/σ in equation (2) is known as the *shape* parameter because the shape of the Weibull density depends on the value of c. Furthermore, the parameter a in equation (3) or $1/m$ in equation (2) is called the *scale* parameter for the Weibull distribution because the effect of different values of a in equation (3) on the graph of the density is just to change the scale on the horizontal x-axis, and not the basic shape of the graph. The parameter b in equation (3) may be described as a *location* or *shift* parameter. In our opinion, the widely-used parametrization of equation (3) is more meaningful and natural.

4 Application to Hong Kong Mortality Tables

To illustrate the applications of the Heligman and Pollard (1980) and Carrier (1992) models, we now apply equations (1), (2), and (4) to several Hong Kong mortality tables. These equations are applied to the *smoothed mortality tables* and not to the raw mortality rates. The male and female tables are fitted separately.

4.1 1993 Hong Kong Assured Lives Mortality Tables

We apply equations (1), (2), and (4) to the 1993 Hong Kong Assured Lives Mortality Tables. Following the examples given in Heligman and Pollard (1980) and Carrier (1992), we estimate the parameters by minimizing the loss function L

$$L = \sum_{x=0}^{99} \left(1 - \frac{\hat{q}_x}{q_x}\right)^2 \tag{5}$$

where $\hat{q}_x$ is the estimate of q_x. This commonly-used loss function is based on the sum of squared relative errors. All parameter estimates and the loss given in Tables 2 and 3 are calculated using SAS, a statistical software package.[1] For equation (2), we use the Inverse-Gompertz for teenage years because it fits both male and female tables better than the Inverse-Weibull. For the 1993 Hong Kong Assured Lives Mortality male table, the parameter values for the Weibull distribution are $m_1 = 719.42446$ and $\sigma_1 = 5956.388042$ in the Carriere model of equation (2). Hence, it is inappropriate to interpret m_1 and σ_1 as location and dispersion parameters, respectively, in this case.

The term GH^x can be expressed as H^{x-x_0} where x_0 is the age at which $q_x/(1-q_x) = 1$ simply because the first and second terms of equations (1) and (4) are extremely small at that age. Admittedly, x_0 is close to the end of the life table.

Table 2
Parameter Estimates Using Equations (1), (2), and (4) for the 1993 Hong Kong Assured Lives Mortality Male Table

Heligman-Pollard Model:		
Equation (1) with Loss = 0.138381		
$A = 0.000474$	$B = 0.000475$	$C = 0.063875$
$D = 0.000221$	$E = 5.76254$	$F = 17.616056$
$G = 0.0000177511$	$H = 1.104655$	
Carriere Model:		
Equation (2) with Loss = 0.372211		
$\omega_1 = 0.023591$	$m_1 = 719.42446$	$\sigma_1 = 5956.388042$
$\omega_2 = 0.004303$	$m_2 = 17.825229$	$\sigma_2 = 6.361696$
$\omega_3 = 0.972106$	$m_3 = 87.200722$	$\sigma_3 = 10.121226$
Modified Heligman-Pollard Model:		
Equation (4) with Loss = 0.145928		
$A = 0.000474$	$B = 0.000484$	$C = 0.063961$
$D = 0.003088$	$E = 31.092184$	$F = 0.820805$
$G = 0.0000176876$	$H = 1.104667$	

In equation (4), x_0 is 109.9267 for males and 114.2788 for females. Based on the same model, comparison can be made between the two mortality tables. For instance, in equation (4), the value of G is higher and the values of x_0 is lower for males than for females, indicating higher male mortality. Actually, the same phenomenon can be seen in the parameter estimates derived from other models.

[1]For more information on SAS see, for example, *SAS/STAT User's Guide*, Version 6.

Table 3
Parameter Estimates Using Equations (1), (2), and (4)
for the 1993 Hong Kong Assured Lives Mortality Female Table

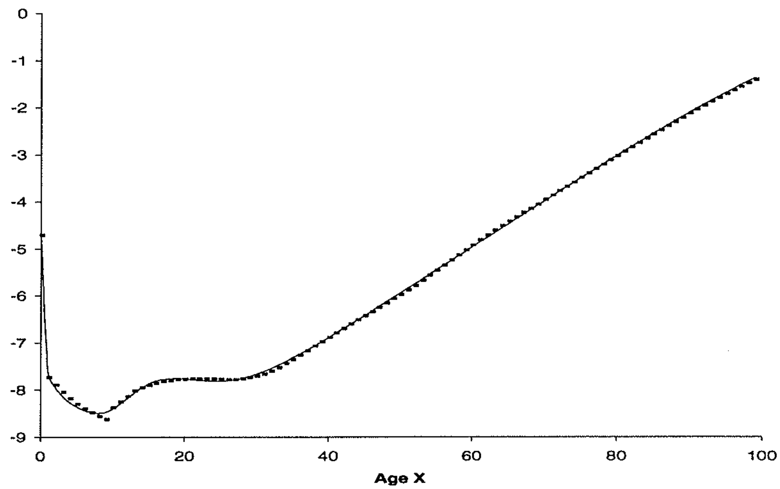
Heligman-Pollard Model:		
Equation (1) with Loss = 0.387583		
$A = 0.000474$	$B = 0.000378$	$C = 0.0622$
$D = 0.000244$	$E = 3.110173$	$F = 22.727553$
$G = 0.0000095144$	$H = 1.106253$	
Carriere Model:		
Equation (2) with Loss = 0.580548		
$\omega_1 = 0.019969$	$m_1 = 105.354377$	$\sigma_1 = 847.670466$
$\omega_2 = 0.008744$	$m_2 = 23.792527$	$\sigma_2 = 11.304304$
$\omega_3 = 0.971287$	$m_3 = 92.370704$	$\sigma_3 = 9.903073$
Modified Heligman-Pollard Model:		
Equation (4) with Loss = 0.418212		
$A = 0.000473$	$B = 0.000392$	$C = 0.062592$
$D = 0.006058$	$E = 13.989076$	$F = 0.892388$
$G = 0.0000091677$	$H = 1.106835$	

The pattern of mortality for the 1993 Hong Kong Assured Lives Mortality male and female tables can be described by the function $\ln(q_x)$ and is displayed in Figures 2 and 3, which plot $\ln(q_x)$ and $\ln(\hat{q}_x)$ for $x = 0, \ldots, 99$ for male and female lives respectively. The male and female tables exhibit similar mortality pattern at childhood and adult ages except that the dip at around age 10 is lower for females.

Tables 2 and 3 show that the estimates of F in equation (1) are 17.6 for males and 22.7 for females. It does appear in Figures 2 and 3 that the hump has its peak at about these ages. The wider spread of the hump for female is reflected by the smaller value of E. The level of mortality in this region for both tables are more or less the same because there is only a slight difference in the values of D.

The estimated mortality rates fit the actual pattern reasonably well for each model. By comparing the loss, the Heligman-Pollard model of equation (1) is slightly better than equation (4). On the other hand, equation (2) does not fit as close as equation (1) and equation (4), especially for the 1993 Hong Kong Assured Lives Mortality male table.

Figure 2
1993 Hong Kong Assured Lives Mortality Male Table
Plot of $\ln(q_x)$ and $\ln(\hat{q}_x)$ Using Equation (1)

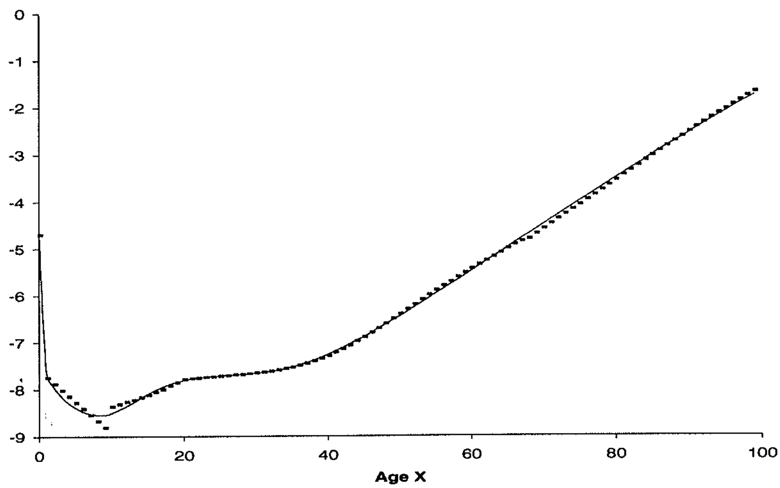


4.2 1991 Hong Kong Female Life Table

In addition, we fit equation (1) and equation (4) to the 1991 Hong Kong Female Life Table. Again, both models fit the pattern of mortality very well. Equation (4) has a smaller loss this time. This fact shows that equation (1) does not always give the best fit among the three models. It depends on the underlying pattern of mortality. The results are given in Tables 4 and 5.

The plots of residuals for Figures 2 and 3 are shown in Figures 4 and 5, respectively. Some of the residuals in the childhood ages are large. Particularly, equation (2) produces relatively large residuals in this region partly because the two parameter Weibull is used instead of the more flexible three parameter Weibull. The plots also indicate the presence of systematic bias at certain ages. This weakness is the price that we have to pay for using a parametric model to fit the entire mortality table that already contains smoothed values.

Figure 3
1993 Hong Kong Assured Lives Mortality Female Table
Plot of $\ln(q_x)$ and $\ln(\hat{q}_x)$ Using Equation (1)



5 Some Closing Remarks

We should be aware that, in Section 4, the equations are fitted to the *smoothed mortality tables*, and not to the raw mortality rates, thereby constituting a rather unusual type of two stage smoothing process. If we fit the equations to the raw data, the loss may become larger than those shown in Tables 2 to 5. Under certain circumstances, the NLIN procedure of SAS is successful in minimizing equation (5) only if it has good starting values. In particular, the Carriere model of equation (2) converges slowly due to its complicated nature, compared to the other two models.

The fit during the childhood years could be improved for equation (2). This point was mentioned in Carriere's original paper as well. To obtain a better fit, we may employ the three parameter Weibull distribution with the idea of truncated survival distribution, instead of using the standard (two parameter) Weibull distribution. This change should provide at least a better interpretation of the parameters in the first term of the model.

Figure 4
1993 Hong Kong Assured Lives Mortality Male Table Residuals Resulting From Using Equation (1)

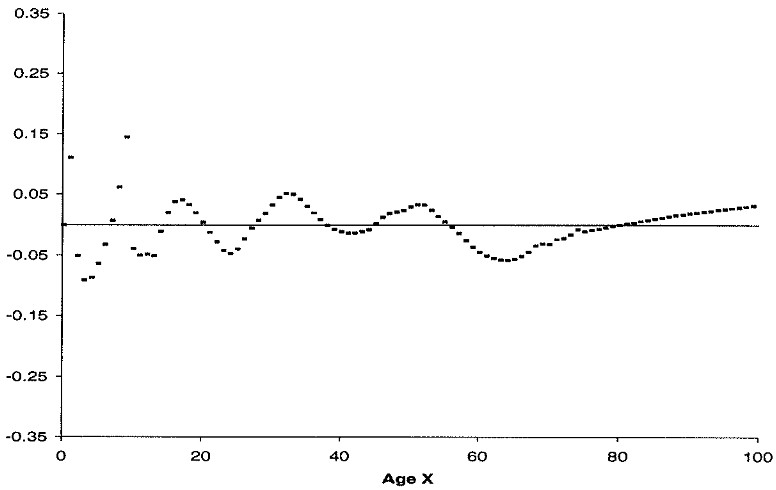


Figure 5
1993 Hong Kong Assured Lives Mortality Female Table Residuals Resulting From Using Equation (1)

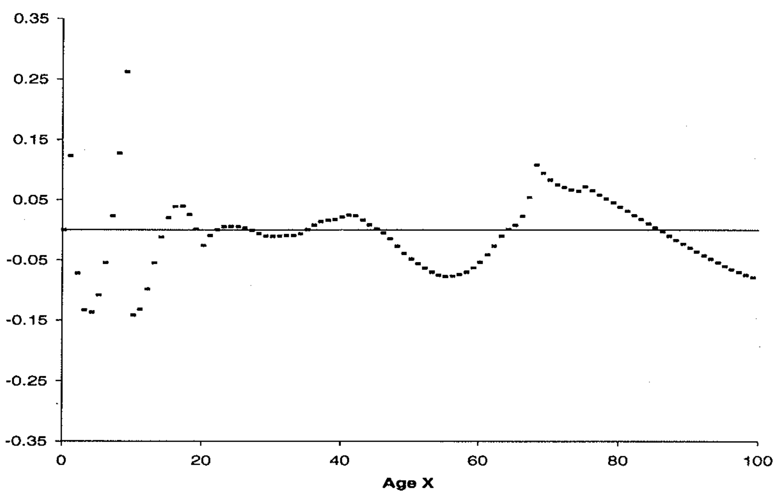


Table 4
Parameter Estimates Using Equation (1)
for the 1991 Hong Kong Female Mortality Table

Heligman-Pollard Model:	
Equation (1) with Loss = 0.14638	
A = 0.000753	E = 0.888422
B = 0.158423	F = 23.71991
C = 0.167371	G = 0.0000076034
D = 0.000226	H = 1.11813

Table 5
Parameter Estimates Using Equation (4)
for the 1991 Hong Kong Female Mortality Table

Modified Heligman-Pollard Model:	
Equation (1) with Loss = 0.122038	
A = 0.001015	E = 4.946691
B = 0.317964	F = 0.933656
C = 0.229451	G = 0.000007981
D = 0.009157	H = 1.117485

Carriere (1994) proposes a select and ultimate parametric model which is based on equation (2). We feel that equation (1) also may be used to construct his select and ultimate parametric model because equation (1) has a much simpler form. This reconstruction seems feasible.

References

Carriere, J.F. "A Select and Ultimate Parametric Model." *Transactions of the Society of Actuaries* 46 (1994): 33–50.

Carriere, J.F. "Parametric Models for Life Tables." *Transactions of the Society of Actuaries* 44 (1992): 77–99.

Forfar, D.O., McCutcheon, J.J. and Wilkie, A.D. "On Graduation by Mathematical Formula." *Journal of the Institute of Actuaries* 115 (1988): 1–149.