TRANSACTIONS OF SOCIETY OF ACTUARIES
1994 VOL. 46

A SELECT AND ULTIMATE PARAMETRIC MODEL

JACQUES F. CARRIERE

ABSTRACT

This paper presents a parsimonious 11-parameter model that explains the pattern of mortality for the female and male mortality rates of the 1975–80 Select and Ultimate Basic Tables. This parametric model is useful because it can predict the select mortality rates beyond the 15-year select period and because it can predict the select rates for issue ages greater than 70 years old. Moreover, the parameters in this model provide insightful statistical information about the data.

1. INTRODUCTION

The purpose of this paper is to present a parsimonious parametric model that explains the pattern of mortality for select and ultimate mortality tables. Specifically, we model the 1975–80 Basic Tables produced by the Committee on Ordinary Insurance and Annuities [2].

A review of the literature reveals that very little research has been done on the fitting of parametric formulas to select rates. Using Canadian data, Panjer and Russo [5] did a graduation of select and ultimate rates that they refer to as "parametric." In fact, a true parametric formula was developed only at the higher ages. This is also true of the laws of select and ultimate mortality developed by Tenenbein and Vanderhoof [8]. In both cases, the formulas are based on Gompertz's law or generalizations thereof, and in neither case were they able to develop formulas that fit the pattern of mortality from childhood to early adulthood. In contrast, this paper presents a parametric formula that reflects the fall in mortality at the childhood years, the hump at about age 20 and the exponential pattern at the adult ages.

We suggest that future graduations be done with mathematical formulas because of the many advantages of this approach. First, the basic tables present select rates for five-year age groupings, which may be inconvenient to practitioners, and so Paquin [6] had to extend the select rates to the issue ages $x=0, 1, \ldots, 70$ while ensuring that a monotonicity constraint, given in Equation (4.2), holds. If the committee [2] had presented the select rates in the form of a mathematical formula, then Paquin's interpolation exercise would not be necessary. Second, another strength of a mathematical formula is its ability to predict or estimate the select rates at issue ages above 70, which is impossible with the current tabulated rates. Still another strength of a parametric model is its ability to extend the select period beyond 15 years. Therefore, a mathematical formula is the most convenient way for practitioners to use select rates.

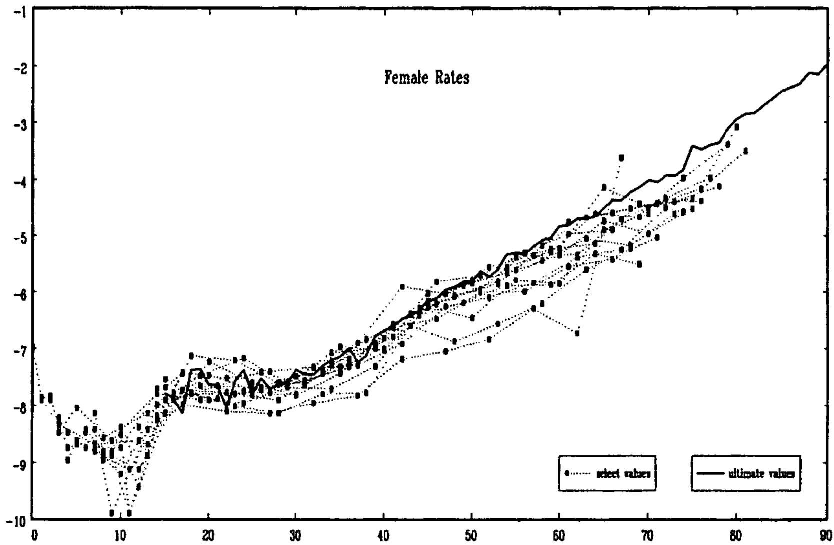
Before we proceed, it is instructive to plot the crude and graduated rates and examine the pattern of mortality in the Basic Tables. All the graphs in this paper were produced with the statistical computing language GAUSS. Figure 1 gives plots of the logarithm of the select and ultimate *crude* rates for the female and male Basic Tables, while Figure 2 gives plots of the logarithm of the *graduated* rates. Let $\hat{q}_{[x]+k}$ denote a crude select mortality rate for an issue age $x \geq 0$, at the nearest birthday, and for policy year $k+1 \geq 1$. We denote the attained age as $y=x+k$. Note that the select period for these tables is 15 years, and so $\hat{q}_{[x]+k}$ is given only for $k=0, \ldots, 14$. Next, let $\hat{q}_y$ denote a crude ultimate rate for a person aged y. Now, consider the female graph in Figure 1. This graph plots $\log_e(\hat{q}_y)$ for $y=15, \ldots, 90$, and it plots 15 curves for each of the policy years. That is, for each k, the graph plots $\log_e(\hat{q}_{[y-k]+k})$ for the attained ages $y=k, \ldots, k+67$. Many of the values for $\hat{q}_{[y-k]+k}$ are not given in the Basic Tables because of grouping; therefore the function $\log_e(\hat{q}_{[y-k]+k})$, with respect to y, was approximated linearly.

Next, let $\tilde{q}_{[x]+k}$ denote a graduated select mortality rate for an issue age $x \geq 0$, at the nearest birthday, and for policy year $k+1 \geq 1$; and let $\tilde{q}_y$ denote a crude ultimate rate for a person aged y. Figure 2 is the same graph as Figure 1 except that we replaced the crude rates, $\hat{q}$, with the graduated rates, $\tilde{q}$. Figure 2 illustrates that the pattern exhibited by the graduated rates shows a decrease in the childhood years, a hump at about age 20 and a linear pattern at the adult ages.

The rest of the paper proceeds as follows. First, we present our mathematical formula and discuss some of its features, including the choice of parametrization. In this discussion, we also present an approximation for the mean and variance of the Gompertz distribution. Next, we examine some issues associated with the lack of information in the data presented by the committee [2]. Finally, we estimate the parameters and discover some good-fitting formulas that have a monotonic property.

FIGURE 1

THE LOGARITHMS OF THE CRUDE RATES FROM THE MALE AND FEMALE 1975–80 BASIC TABLES
THE HORIZONTAL AXIS GIVES THE ATTAINED AGE y,
WHILE THE VERTICAL AXIS GIVES THE VALUES $\log_e(\hat{q}_y)$ AND $\log_e(\hat{q}_{[y-k]+k})$.



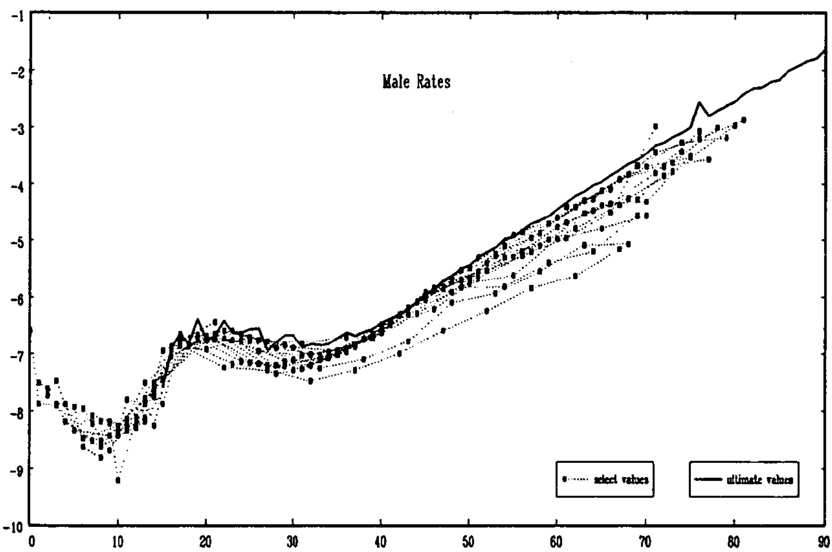
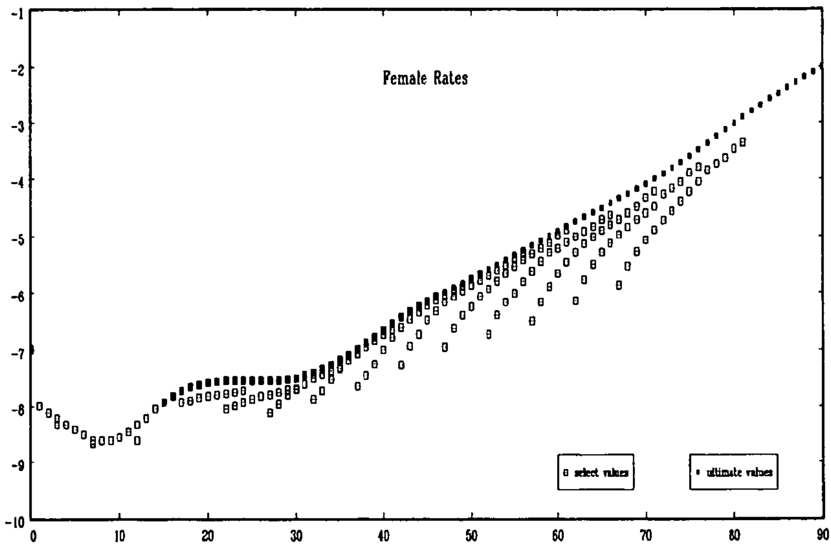
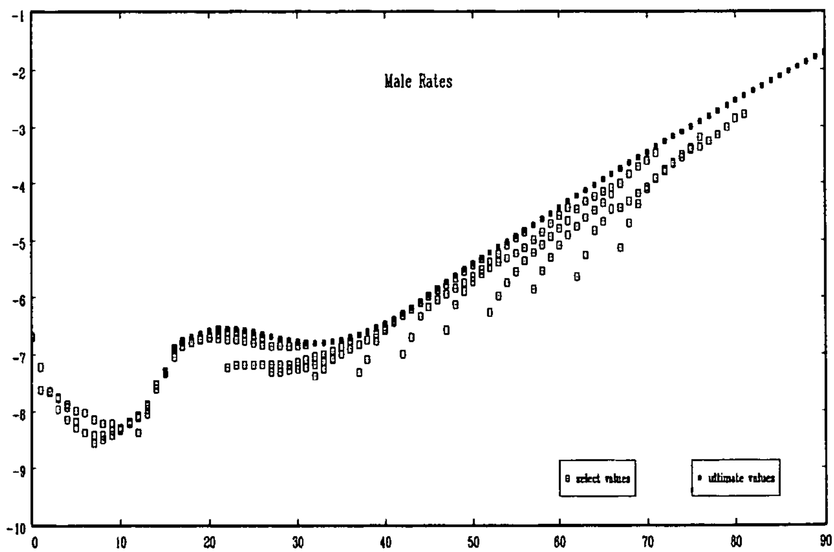


FIGURE 2

THE LOGARITHMS OF THE GRADUATED RATES
FROM THE MALE AND FEMALE 1975–80 BASIC TABLES
THE HORIZONTAL AXIS GIVES THE ATTAINED AGE y,
WHILE THE VERTICAL AXIS GIVES THE VALUES $\log_e(\bar{q}_y)$ AND $\log_e(\bar{q}_{[y-k]+k})$.





2. A PARAMETRIC MODEL

In this section, we present our mathematical law of select and ultimate mortality and discuss some of its features. A general mathematical law of select and ultimate mortality can be defined as follows. Let $x \geq 0$ be the issue age at the nearest birthday; let $k+1 \geq 1$ be the policy year; and let $y=x+k$ be the attained age. Also, let $s(y|\boldsymbol{\theta}_k)$ denote a parametric survival function with a parameter vector, $\boldsymbol{\theta}_k$, that converges to the finite value $\boldsymbol{\theta}_\infty$ as $k \rightarrow \infty$. Then, the select mortality rates can be defined as

$$q_{[y-k]+k} = 1 - \frac{s(y+1|\boldsymbol{\theta}_k)}{s(y|\boldsymbol{\theta}_k)}, \tag{2.1}$$

while the ultimate rates can be defined as

$$q_y = 1 - \frac{s(y+1|\boldsymbol{\theta}_\infty)}{s(y|\boldsymbol{\theta}_\infty)}. \tag{2.2}$$

Now, let us specify the formula for $s(y|\boldsymbol{\theta})$. The pattern of mortality exhibited in Figure 2 is very similar to that of the total population of the U.S. A model developed by Carriere [1] proved successful in modeling the pattern of mortality of the U.S. population. Therefore, we propose to use Carriere's model as the basic formula $s(y|\boldsymbol{\theta})$ in (2.1) and (2.2). The model given by Carriere [1] is a mixture of a Weibull survival function, an Inverse-Weibull survival function and a Gompertz survival function. In this eight-parameter model, the probability of surviving to age $y>0$ is

$$s(y|\boldsymbol{\theta}) = \psi_1 s_1(y) + \psi_2 s_2(y) + \psi_3 s_3(y) \tag{2.3a}$$

where

$$s_1(y) = \exp\left\{-\left(\frac{y}{m_1}\right)^{m_1/\sigma_1}\right\}, \tag{2.3b}$$

$$s_2(y) = 1 - \exp\left\{-\left(\frac{y}{m_2}\right)^{-m_2/\sigma_2}\right\}, \tag{2.3c}$$

$$s_3(y) = \exp\left\{e^{-m_3/\sigma_3} - e^{(y-m_3)/\sigma_3}\right\}, \tag{2.3d}$$

$$\psi_3 = 1 - \psi_1 - \psi_2, \tag{2.3e}$$

$$\boldsymbol{\theta} = (\psi_1, \psi_2, \psi_3, m_1, m_2, m_3, \sigma_1, \sigma_2, \sigma_3)'. \tag{2.3f}$$

The parameters in this model are $\psi_i \in [0, 1]$, $m_i > 0$ and $\sigma_i > 0$ for $i=1$, 2, 3, and they are summarized with the vector $\boldsymbol{\theta}$. It seems that we have 9 parameters, but there are only 8 because of the restriction $\psi_3 = 1 - \psi_1 - \psi_2$. This nonstandard parametrization of the Gompertz, Weibull and Inverse-Weibull functions was chosen because it provides insightful statistical information. For example, m_i is approximately equal to the mean of $s_i(y)$, while σ_i is proportional to the standard deviation of $s_i(y)$. First, it is easy to verify that all the mass concentrates about m_i when σ_i is small. This is true because

$$\lim_{\sigma_i \to 0} \{s_i(m_i - \epsilon) - s_i(m_i + \epsilon)\} = 1,$$

for any arbitrary $\epsilon > 0$. Let μ_i and υ_i denote the mean and variance of the survival function $s_i(y)$ for $i=1$, 2, 3. Consulting Johnson and Kotz [3], we find that the Weibull distribution admits the approximations

$$\mu_1 \approx m_1 - \gamma\sigma_1 \quad \text{and} \quad \upsilon_1 \approx \frac{\sigma_1^2 \pi^2}{6},$$

whenever σ_1 is small. The value $\gamma = 0.5772 \ldots$ in the approximation of μ_1 is Euler's constant. Using this result for the Weibull, we can show that for the Inverse-Weibull, $\mu_2 \approx m_2 + \gamma\sigma_2$ and $\upsilon_2 \approx \sigma_2^2\pi^2/6$, when σ_2 is small. Finally, the Gompertz distribution is a truncated extreme-value distribution. Consulting Johnson and Kotz [3], we find that the mean of the extreme-value distribution

$$1 - \exp\{-e^{(y-m)/\sigma}\}$$

is $m - \gamma\sigma$ and the variance is $\sigma^2\pi^2/6$. If $m_3 > 0$, then we can show that for the Gompertz distribution

$$\mu_3 \approx m_3 - \gamma\sigma_3 \quad \text{and} \quad \upsilon_3 \approx \frac{\sigma_3^2 \pi^2}{6},$$

when σ_3 is small. We can also prove that m_3 is the mode of the probability density function of the Gompertz survival function $s_3(y)$. Therefore, we can conclude that $m_i \approx \mu_i$ and $1.28 \times \sigma_i$ is approximately equal to the standard deviation of $s_i(y)$ when σ_i is small. All the approximations were verified numerically, with good success.

Now, let us model $\boldsymbol{\theta}_k = (\psi_{1,k}, \psi_{2,k}, \psi_{3,k}, m_{1,k}, m_{2,k}, m_{3,k}, \sigma_{1,k}, \sigma_{2,k}, \sigma_{3,k})'$ for $k \geq 0$. With this notation, $\boldsymbol{\theta}_0$ denotes the parameter values at issue, while $\boldsymbol{\theta}_\infty=(\psi_{1,\infty}, \psi_{2,\infty}, \psi_{3,\infty}, m_{1,\infty}, m_{2,\infty}, m_{3,\infty}, \sigma_{1,\infty}, \sigma_{2,\infty}, \sigma_{3,\infty})'$ denotes the ultimate parameter values. Using $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\infty$, we defined $\boldsymbol{\theta}_k$ as a weighted average

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_0 + (\boldsymbol{\theta}_\infty - \boldsymbol{\theta}_0)\,(1 - \exp\{-ak^b\}), \tag{2.4}$$

where $a>0$ and $b>0$. An idea similar to (2.4) was used by Panjer and Giuseppe [5], where a weighted average was taken of q_y and $q_{[y]}$. Note that if $k\rightarrow\infty$, then $\boldsymbol{\theta}_k\rightarrow\boldsymbol{\theta}_\infty$, as required.

If we use (2.4), then the resulting model will have 18 parameters. Other formulas for $\boldsymbol{\theta}_k$, with more than 18 parameters, can be constructed, but we found that (2.4) yielded a good-fitting model of the 1975–80 basic rates. Actually, we found that restricting the parameters as follows

$$\begin{gathered}\psi_{1,0} = \psi_{1,\infty}, \quad \psi_{2,0} = \psi_{2,\infty},\\ m_{1,0} = m_{1,\infty}, \quad m_{2,0} = m_{2,\infty},\\ \sigma_{1,0} = \sigma_{1,\infty},\end{gathered} \tag{2.5}$$

yielded a 13-parameter model that also fit the data well. By adding more restrictions, we discovered some parsimonious models that had less than 13 parameters that also fit the data. See Section 4 for more details.

3. HETEROSCEDASTICITY AND THE LOSS FUNCTION

In this section, we discuss the issue of heteroscedasticity (heterogeneous variance) and the related issue of choosing a loss function for parameter estimation. Suppose we have crude rates $\hat{q}_x$ where $x\in\mathbf{X}$ and we want to model the response with a parametric function $q_x(\boldsymbol{\theta})$. In a nonlinear regression model, we would assume that $\hat{q}_x=q_x(\boldsymbol{\theta})+\epsilon_x$ where $E(\epsilon_x)=0$. In our case, the variance $\mathrm{Var}(\hat{q}_x)=\sigma_x^2$ is not constant in x (heteroscedastic), and so Seber and Wild [7] would suggest that we estimate the parameters by the method of weighted least-squares, to avail ourselves of some standard inference results. Assuming that $\hat{q}_x$ is uncorrelated with $\hat{q}_y$ when $x\neq y$, Seber and Wild [7] suggest that we minimize the *loss function*

$$\sum_{x\in\mathbf{X}} w_x\{\hat{q}_x - q_x(\boldsymbol{\theta})\}^2, \tag{3.1}$$

where the weights w_x are proportional to the inverse of the variance, $1/\sigma_x^2$. Let us derive an expression for the variance σ_x^2. Many of the ideas in the following discussion can be found in Klugman [4].

Let n_x be the number of policies associated with the crude rate $\hat{q}_x$ and let D_x be the total amount of death claims associated with $\hat{q}_x$. The amount of death claims, D_x, is based on a group of n_x policies, where the death benefit for policy $i=1, \ldots, n_x$ is $b_{x,i}$. Let $\delta_{x,i}$ denote an indicator random variable that is equal to 1 if a death has occurred and 0 otherwise. Assume that $\delta_{x,1}, \ldots, \delta_{x,n_x}$ are independent and identically distributed with $E(\delta_{x,i})=q_x$. Then, $\hat{q}_x=D_x/B_x$, where

$$D_x = \sum_{i=1}^{n_x} b_{x,i}\delta_{x,i}$$

and

$$B_x = \sum_{i=1}^{n_x} b_{x,i}.$$

This means that $E(\hat{q}_x)=q_x$, and so the crude rate is an unbiased estimator of q_x. We can now calculate the variance of $\hat{q}_x$, given that we know $b_{x,i}$ for $i=1, \ldots, n_x$. This is equal to

$$\sigma_x^2 = \frac{q_x(1 - q_x) \times \sum_{i=1}^{n_x} b_{x,i}^2}{B_x^2} \tag{3.2}$$

The variance σ_x^2 is not calculable because

$$\sum_{i=1}^{n_x} b_{x,i}^2$$

is not given in [2]. Let us make some simplifying assumptions. Specifically, let us assume that

$$n_x^{-1} \sum_{i=1}^{n_x} b_{x,i} = \alpha$$

and

$$n_x^{-1} \sum_{i=1}^{n_x} b_{x,i}^2 = \beta$$

are constant functions in x. With these assumptions, we find that

$$\sigma_x^2 = \frac{(\beta/\alpha^2) \times q_x(1 - q_x)}{n_x}.$$

We know that $\hat{q}_x = D_x/(n_x\alpha)$, where D_x is the amount of death claims associated with $\hat{q}_x$. Therefore $E(D_x) = q_x n_x \alpha$ and

$$\sigma_x^2 = \frac{(\beta/\alpha)(1 - q_x)q_x^2}{E(D_x)} \approx \frac{(\beta/\alpha)q_x^2}{E(D_x)} \approx \frac{(\beta/\alpha)\hat{q}_x^2}{D_x}.$$

So, if we let $w_x = D_x/\hat{q}_x^2$, then the weights will be approximately proportional to the inverse of the variance and the *residuals*

$$\sqrt{D_x}\left\{1 - \frac{q_x(\boldsymbol{\theta})}{\hat{q}_x}\right\}, x \in \mathbf{X}, \tag{3.3}$$

will almost have a constant variance, as required.

Initially, we used a loss function like

$$\sum D_x \left\{1 - \frac{q_x(\boldsymbol{\theta})}{\hat{q}_x}\right\}^2$$

for estimating the parameters, but we found that the tail of the distribution of residuals was heavy because of many outliers. We believe that these outliers are due to a violation of our initial assumption that the variance of $\hat{q}_x$ is always proportional to $D_x/\hat{q}_x^2$. Nevertheless, we believe that letting $w_x = D_x/\hat{q}_x^2$ removes most of the heteroscedasticity in the residuals. As a precaution, we decided to use a loss function with absolute residuals to reduce the influence from outliers, as recommended by Seber and Wild [7]. In conclusion, all the parameter estimates in this paper can be found by minimizing loss functions that have the form

$$\sum \sqrt{D_x}\left|1 - \frac{q_x(\boldsymbol{\theta})}{\hat{q}_x}\right|. \tag{3.4}$$

This approach to parameter estimation is different than any of the methods proposed by Carriere [1] and Tenenbein and Vanderhoof [8]. We found that (3.4) leads to reasonable parameter estimates.

4. PARAMETER ESTIMATION

In this section, we estimate the parameters a, b, $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\infty$ that yield a good fit to the male and female select and ultimate crude rates given in the 1975–80 Basic Tables. All parameter estimates were calculated by the NONLIN module of the statistical computer software called SYSTAT. We found that the NONLIN *simplex* or polytope algorithm was very successful at minimizing the nondifferentiable loss function

$$L(a, b, \boldsymbol{\theta}_0, \boldsymbol{\theta}_\infty) = \sum_{x \in \mathbf{X}} \sum_{k=0}^{14} w_{[x]+k} \left| 1 - \frac{q_{[x]+k}(a, b, \boldsymbol{\theta}_0, \boldsymbol{\theta}_\infty)}{\hat{q}_{[x]+k}} \right| + \sum_{y=15}^{100} w_y \left| 1 - \frac{q_{[y-24]+24}(a, b, \boldsymbol{\theta}_0, \boldsymbol{\theta}_\infty)}{\hat{q}_y} \right|, \quad (4.1)$$

which is a generalization of (3.4). If you look at (2.1), then you will find that (4.1) is well-defined when $y=15$. The simplex algorithm is successful in minimizing (4.1) only if it has good starting values. We started with values given in Carriere [1], but in the future we would use the parameter estimates developed in this paper.

In our loss function, $\mathbf{X}=\{0, 1, 3, 7, 12, 17, 22, \ldots, 67\}$ and $q_{[x]+k}(a, b, \boldsymbol{\theta}_0, \boldsymbol{\theta}_\infty)$ is our parametric formula for a select mortality rate when the issue age is x and the policy year is $k+1$. The value $\hat{q}_{[x]+k}$ denotes a crude select rate, while $\hat{q}_y$ denotes a crude ultimate rate for a person aged y. Let $D_{[x]+k}$ denote the amount of death claims associated with $\hat{q}_{[x]+k}$ and let D_y denote the amount of death claims associated with $\hat{q}_y$. Define

$$\Sigma = \sum_{x \in \mathbf{X}} \sum_{k=0}^{14} \sqrt{D_{[x]+k}} + \sum_{y=15}^{100} \sqrt{D_y}.$$

Then the weights are equal to

$$w_{[x]+k} = \frac{\sqrt{D_{[x]+k}}}{\Sigma} \quad \text{and} \quad w_y = \frac{\sqrt{D_y}}{\Sigma}.$$

This definition for our weights allows us to interpret the loss function $L(\cdot)$ as an average of absolute relative errors, giving us a meaningful comparison of the loss for the female and male models that we develop. This is necessary because the amount of death claims from the male experience is about ten times that of the female experience.

Let us further justify the form of the loss function given in (4.1). This loss function uses all the crude data given in [2], except for the select rates for 70 and over, as a group. We excluded these rates because we were unable to determine the appropriate issue age for this group. Also included in (4.1) are the ultimate crude rates that are based on the experience from policy years 16 and over. This means that the ultimate crude rate, $\hat{q}_y$, actually corresponds to a parametric select rate with an average policy year of about $k+1=25$. The predicted value of 24, in the expression $q_{[y-24]+24}(a, b, \boldsymbol{\theta}_0, \boldsymbol{\theta}_\infty)$, was chosen after some preliminary analysis in which we predicted the graduated ultimate rates with a parametric model that was constructed with the select graduated rates only.

Using SYSTAT, we were able to find parameter values a, b, $\boldsymbol{\theta}_0$, and $\boldsymbol{\theta}_\infty$ that minimized (4.1) for the female and male rates separately. Tables 1 and 2 give the parameter estimates for the female and male models, respectively. Note that the tables do not include estimates for the parameters $\psi_{1,\infty}$, $\psi_{2,\infty}$, $m_{1,\infty}$, $m_{2,\infty}$, and $\sigma_{1,\infty}$, because we used the restrictions of (2.5). We found that introducing these parameters, by removing the restrictions, did not improve the fit very much. For example, we found that the full 18-parameter model had an *average relative error*, $L(\cdot)$, of 0.079, which is a minimal improvement to the 13-parameter model where $L(\cdot)=0.082$. Initially, an 8-parameter model that did not account for the effects of selection was fit to the data. This special case occurs by imposing the constraint $\boldsymbol{\theta}_0=\boldsymbol{\theta}_\infty$. This reduced model explained most of the pattern of mortality in the data because the average relative error was equal to 0.296 and 0.212 for the female and male models, respectively.

The Gompertz component of the 8-parameter model explained most of the deaths because $\psi_{3,0}=1-\psi_{1,0}-\psi_{2,0}$ was equal to 0.99314 and 0.98177 for the female and male models, respectively. This means that the most important parameters are the Gompertz parameters. By removing the restriction $m_{3,0}=m_{3,\infty}$ and by freeing the parameter a while fixing $b=1$, we found a 10-parameter model that improved the fit considerably. Specifically, $L(\cdot)$ reduced to 0.188 and 0.096 for the female and male models, respectively. Removing the restriction $\sigma_{3,0}=\sigma_{3,\infty}$ yielded an 11-parameter model that further improved the fit. At this point, we discovered that adding more parameters to the female model decreased $L(\cdot)$ only minimally. But freeing the parameter b resulted in a 12-parameter male model that was somewhat better. In any case, adding more parameters to the 13-parameter model yielded only minimal decreases in $L(\cdot)$.

TABLE 1

PARAMETER ESTIMATES FOR THE FEMALE MODEL

Model	Eight	Ten	Eleven	Twelve	Thirteen
$\psi_{1,0}$	0.00372	0.00374	0.00335	0.00332	0.00314
$\psi_{2,0}$	0.00314	0.00333	0.00271	0.00246	0.00302
$m_{1,0}$	8.386	9.008	7.638	7.673	6.759
$m_{2,0}$	18.16	18.21	18.72	18.59	18.69
$m_{3,0}$	89.95	102.0	114.2	120.0	119.2
$m_{3,\infty}$	89.95	88.94	88.08	87.76	87.69
$\sigma_{1,0}$	14.00	15.56	13.21	13.35	11.65
$\sigma_{2,0}$	4.384	4.562	4.425	4.405	8.398
$\sigma_{2,\infty}$	4.384	4.562	4.425	4.405	2.573
$\sigma_{3,0}$	10.78	11.86	15.36	16.59	16.57
$\sigma_{3,\infty}$	10.78	11.86	11.25	11.27	11.15
a	0	0.1592	0.1989	0.3227	0.2856
b	1	1	1	0.7873	0.8307
$L(\cdot)$	0.296	0.188	0.172	0.171	0.168
Change		0.108	0.016	0.001	0.003

TABLE 2

PARAMETER ESTIMATES FOR THE MALE MODEL

Model	Eight	Ten	Eleven	Twelve	Thirteen
$\psi_{1,0}$	0.00623	0.00963	0.00941	0.00804	0.00832
$\psi_{2,0}$	0.01200	0.01234	0.01187	0.01046	0.01006
$m_{1,0}$	9.514	30.13	27.55	31.12	23.79
$m_{2,0}$	19.87	20.27	20.05	19.57	19.36
$m_{3,0}$	83.22	92.64	94.37	105.8	105.7
$m_{3,\infty}$	83.22	81.58	81.64	80.18	80.25
$\sigma_{1,0}$	15.28	50.02	49.20	62.56	50.06
$\sigma_{2,0}$	4.711	4.875	4.757	4.635	4.591
$\sigma_{2,\infty}$	4.711	4.875	4.757	4.635	3.641
$\sigma_{3,0}$	9.839	10.48	11.15	14.61	14.56
$\sigma_{3,\infty}$	9.839	10.48	10.46	9.959	9.984
a	0	0.1253	0.1307	0.3684	0.3767
b	1	1	1	0.6136	0.6092
$L(\cdot)$	0.212	0.096	0.091	0.083	0.082
Change		0.116	0.005	0.008	0.001

The values $L(\cdot)$ seem to indicate that the 11-parameter female model and the 12-parameter male model were good-fitting models. But the 12-parameter model for both the male and female data did not pass a monotinicity test. Happily, we found that both the female and male 11-parameter models satisfied the following *monotonicity constraint*

$$q_{[x]+k}(\cdot) \leq q_{[x-1]+k+1}(\cdot), \tag{4.2}$$

for all $x=1$ ..., 78 and $k=0$, ..., $\min(y-1, 25)$, except for the violation $q_{[1]}>q_{[0]+1}$ in the female rates and the violation $q_{[78]+24}>q_{[77]+25}$ in the male rates. We also discovered that the constraint in (4.2) held at issue ages greater than 78 for the female model. The failure in monotonicity of our model at attained ages that are greater than $102=78+24$ is not very significant because no data were available beyond age 100.

Notwithstanding the low values for $L(\cdot)$ and the monotonic property of the 11-parameter model, the most important way of verifying that the parameters for this model actually fit the data is to plot the estimated rates against the crude rates given in the 1975–80 Basic Tables. Figures 3 and 4 are plots of the select rates for the female and male models, respectively. Specifically, each plot shows 15 graphs, one for each $k=0$, ..., 14, of

$$\log_e(\hat{q}_{[y-k]+k}) \text{ at } y \in k + \mathbf{X}$$

and of

$$\log_e(q_{[y-k]+k}(a, b, \boldsymbol{\theta}_0, \boldsymbol{\theta}_\infty)) \text{ at } y = k, \ldots, k + 67.$$

After examining Figures 3 and 4, we believe that the rates calculated with our formulas are almost indistinguishable from the crude select rates in the 1975–80 tables.

Figure 5 gives two graphs, one showing the female ultimate data and the other showing the male ultimate data. Examining these graphs, we find that our 11-parameter models reproduced the pattern of mortality very well. In conclusion, the graphical evidence along with the monotonicity property and the low values for $L(\cdot)$ suggest that our 11-parameter select and ultimate parametric models did a good job.

5. CONCLUSION

In conclusion, Figure 6 illustrates our 11-parameter female and male models at various policy years. This illustration immediately shows that the effects of selection are minimal at the younger ages and that these effects increase at the older ages.

Based on the success of our mathematical law of select and ultimate mortality, in capturing the pattern of mortality in the 1975–80 Basic Tables, we suggest that future graduations be done with parametric models. One advantage of this approach is that the mathematical formula provides a ready extrapolation for issue ages beyond 70. Another advantage is that we can easily extend the select period beyond 15 years. Finally,

FIGURE 3

THE LOGARITHMS OF THE RATES FROM THE CRUDE DATA AND THE 11-PARAMETER FEMALE MODEL

THE HORIZONTAL AXIS GIVES THE ATTAINED AGE y, WHILE THE VERTICAL AXIS GIVES THE VALUES OF $\log_e(\hat{q}_{[y-k]+k})$ AND $\log_e[q_{[y-k]+k}(a, b, \theta_0, \theta_\infty)]$.

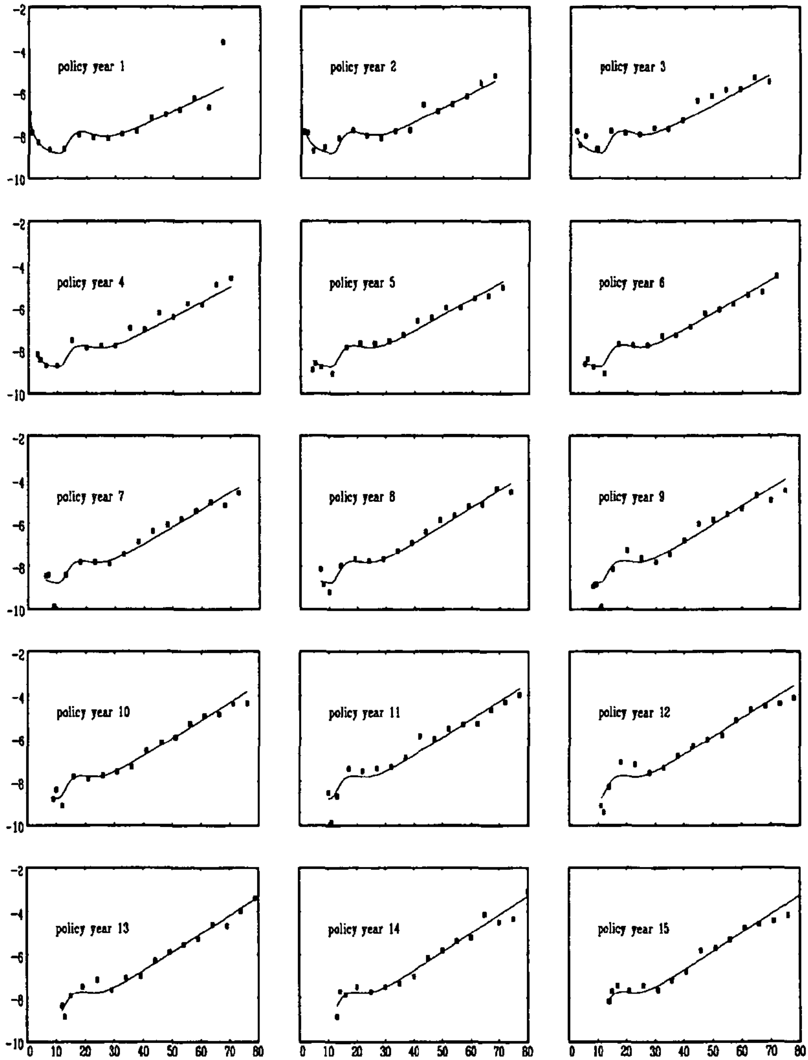


FIGURE 4

THE LOGARITHMS OF THE RATES FROM THE CRUDE DATA AND THE 11-PARAMETER MALE MODEL
THE HORIZONTAL AXIS GIVES THE ATTAINED AGE y, WHILE THE VERTICAL AXIS
GIVES THE VALUES OF $\log_e(\hat{q}_{y|y-k|+k})$ AND $\log_e[q_{|y-k|+k}(a, b, \theta_0, \theta_\infty)]$.

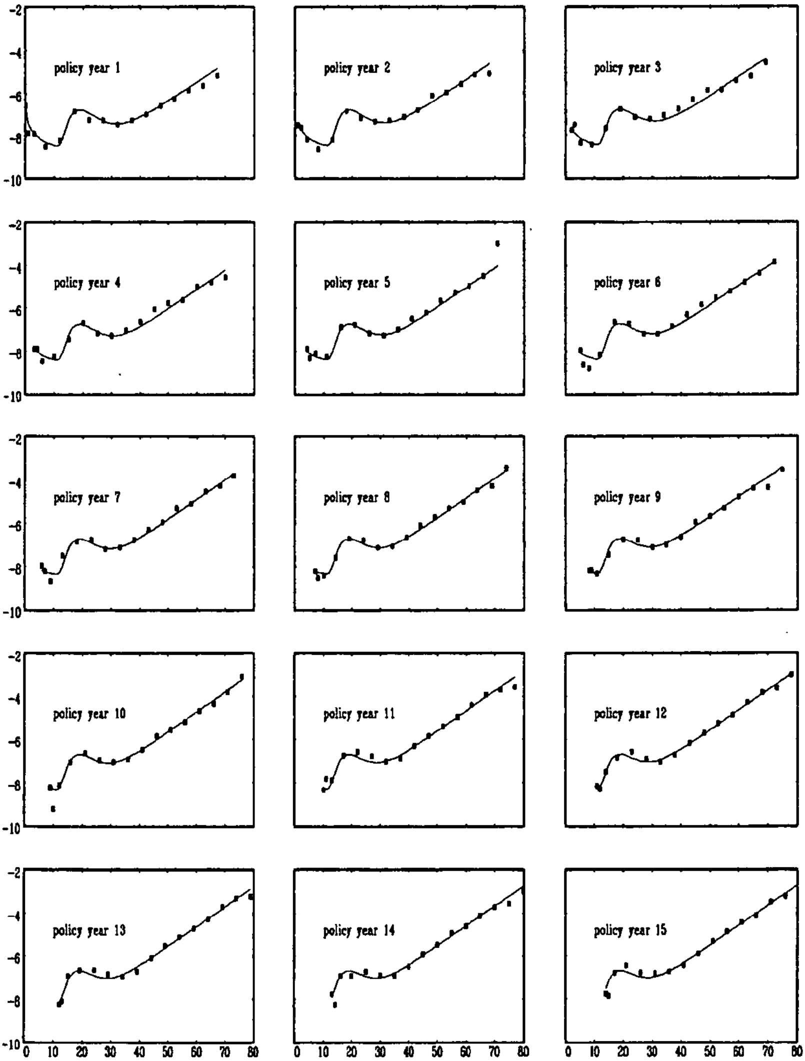


FIGURE 5

A COMPARISON OF THE CRUDE ULTIMATE RATES WITH THE RATES OF THE 11-PARAMETER MODEL
THE HORIZONTAL AXIS GIVES THE ATTAINED AGE y, WHILE THE VERTICAL AXIS
GIVES THE VALUES OF $\log_e(\hat{q}_y)$ AND $\log_e[q_{|y-24|+24}(a, b, \theta_0, \theta_\infty)]$.

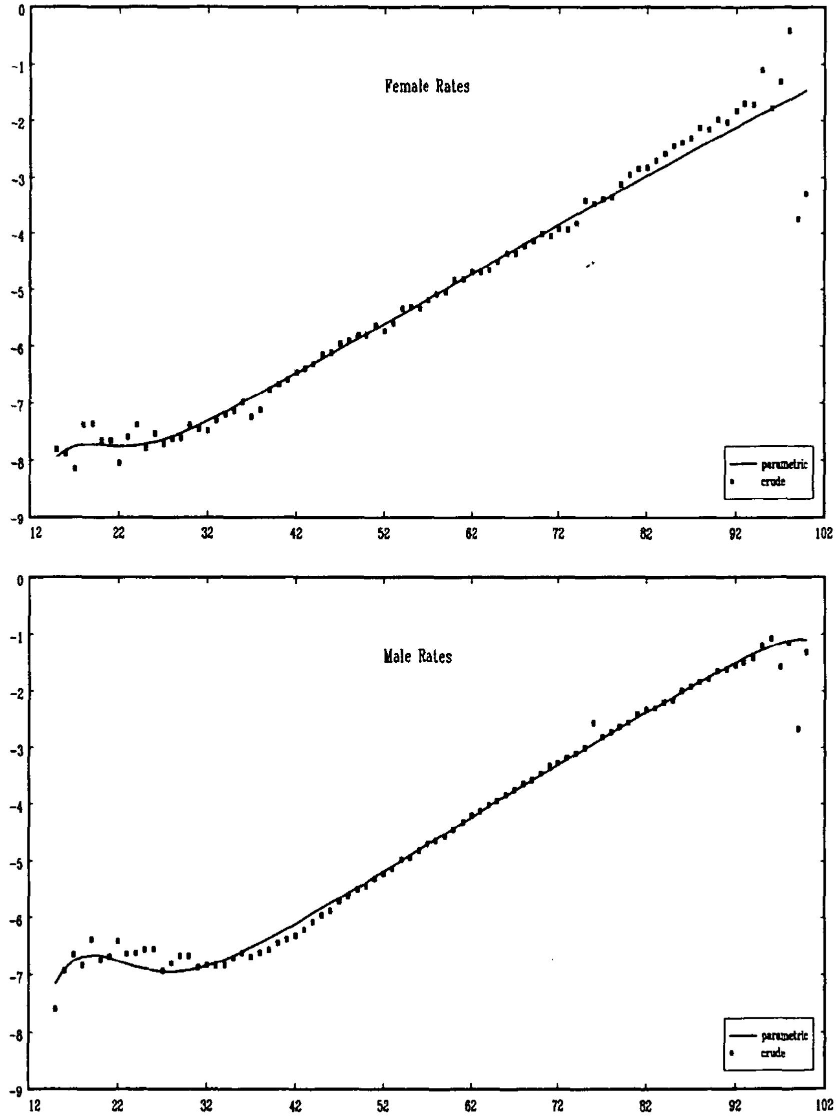
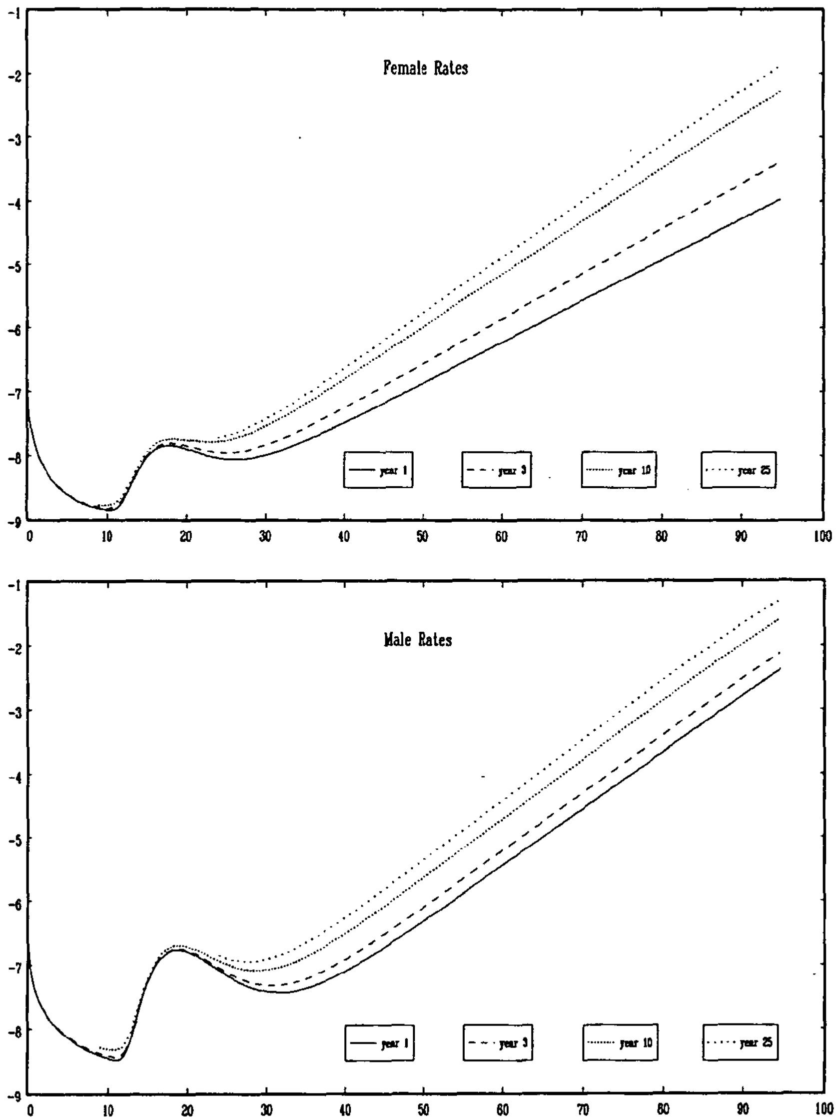


FIGURE 6

LOGARITHMS OF SELECT RATES IN POLICY YEARS $k+1=1$, 3, 10, 25 USING OUR FORMULA
THE HORIZONTAL AXIS GIVES THE ATTAINED AGE y, WHILE THE VERTICAL AXIS
GIVES THE VALUES OF $\log_e[q_{[y-k]+k}(a, b, \theta_0, \theta_\infty)]$.



the parameters in the model provide insightful statistical information about the select rates. Therefore, a mathematical model is the most convenient way for practitioners to calculate select rates.

REFERENCES

1. CARRIERE, J.F. "Parametric Models for Life Tables," *TSA* XLIV (1992): 77–99.
2. COMMITEE ON ORDINARY INSURANCE AND ANNUITIES. "1975-80 Basic Tables," *TSA 1982 Reports* (1982): 55–81.
3. JOHNSON, N.L., AND KOTZ, S. *Distributions in Statistics, 1. Continuous Univariate Distributions*. New York, N.Y.: Wiley, 1970.
4. KLUGMAN, S.A. "On the Variance and Mean Squared Error of Decrement Estimators," *TSA* XXXIII (1981): 301–11.
5. PANJER, H., AND RUSSO, G. "Parametric Graduation of Canadian Individual Mortality Experience: 1982–1988." Waterloo, Ont.: U. of Waterloo, Institute of Insurance and Pension Research, 1990.
6. PAQUIN, C.Y. "An Extension of the 1975–80 Basic Select and Ultimate Mortality Tables, Male and Female—Actuarial Note," *TSA* XXXVIII (1986): 205–27.
7. SEBER, G.A.F, AND WILD, C.J. *Nonlinear Regression*. New York, N.Y.: Wiley, 1989.
8. TENENBEIN, A., AND VANDERHOOF, I.T. "New Mathematical Laws of Select and Ultimate Mortality," *TSA* XXXII (1980): 119–83.

DISCUSSION OF PRECEDING PAPER

MARK D.J. EVANS:

Dr. Carriere has presented an interesting approach to graduating select and ultimate mortality data. He presents comparisons of crude and graduated mortality rates in graphical form with a logarithmic vertical scale. Visually the crude rates and graduation curve appear very similar, but logarithmic scales understate differences when used in this fashion.

For example, consider the numerical data underlying the female rates in Figure 5. These are shown in Table 1 along with the original graduation of the 1975–80 Female Ultimate Mortality Rates for attained ages 79 through 99.

TABLE 1

Attained Age	Crude Mortality Rates	Graduated Mortality Rates		Ratios	
		Carriere	Original	Carriere	Original
79	44.42	39.82	44.00	90	99
80	52.65	43.43	49.48	82	94
81	58.32	47.36	55.51	81	95
82	59.32	51.63	62.09	87	105
83	67.56	56.28	69.22	83	102
84	76.06	61.33	76.90	81	101
85	87.32	66.82	85.13	77	97
86	92.53	72.78	93.91	79	101
87	99.44	79.25	103.24	80	104
88	120.03	86.26	113.12	72	94
89	116.78	93.87	123.55	80	106
90	138.88	102.11	134.53	74	97
91	133.07	111.02	146.06	83	110
92	161.39	120.66	158.14	75	98
93	184.42	131.07	170.77	71	93
94	180.13	142.31	183.95	79	102
95	333.07	154.41	197.68	46	59
96	167.97	167.45	211.96	100	126
97	268.80	181.45	226.79	68	84
98	663.16	196.49	242.17	30	37
99	23.61	212.59	258.10	900	1,093

Dr. Carriere's graduation technique consistently understates the crude data by 10 percent to 25 percent at these ages. The data for the male rates in Figure 5 exhibit a similar problem (but in the opposite direction) in the 40s, as shown in Table 2.

TABLE 2

Attained Age	Crude Mortality Rates	Graduated Mortality Rates		Ratios	
		Carriere	Original	Carriere	Original
36	1.22	1.38	1.34	113	110
37	1.28	1.48	1.26	116	98
38	1.36	1.61	1.35	118	99
39	1.45	1.74	1.42	120	98
40	1.56	1.90	1.60	122	103
41	1.70	2.07	1.72	122	101
42	1.87	2.25	1.84	120	98
43	2.07	2.46	2.02	119	98
44	2.31	2.70	2.30	117	100
45	2.58	2.95	2.59	114	100
46	2.89	3.23	2.80	112	97
47	3.24	3.55	3.33	110	103
48	3.61	3.89	3.63	108	101
49	4.02	4.27	4.13	106	103
50	4.45	4.68	4.37	105	98
51	4.92	5.14	5.00	104	102
52	5.44	5.65	5.42	104	100
53	6.00	6.20	5.91	103	99

These problems with fit would be excessive in practice. Hopefully, refinements of this formula can lead to more useful results.

ROGER SCOTT LUMSDEN*:

Dr. Carriere has written an interesting and timely paper—interesting because there are few examples of parametric fitting to select and ultimate rates and timely because several actuarial experience bodies are currently developing new select and ultimate tables.

I'd like to make a few comments on the weighting factor used in the loss function in Formula (3.1) in Section 3, Heteroscedasticity and the Loss Function.

I have had several opportunities in the last few years to try to develop select and ultimate mortality tables from fairly detailed experience, working on extensions of two-dimensional Whittaker-Henderson graduation suggested in the Knorr paper [2]. And this causes me concern about the weighting factor suggested, which is deaths divided by the square of

*Mr. Lumsden, not a member of the Society, is Actuarial Systems Director, Corporate Actuarial, at Crown Life Insurance Company, Regina, Saskatchewan.

experience mortality rate; an equivalent expression would be the exposures divided by the experience mortality rate. In the largest study (data loaned to me by a large company on condition its name not be disclosed), the exposures and deaths were available for issue ages 0–85 and for durations 1–15 plus ultimate, without grouping. For males the total deaths were \$1.8 billion and for females \$0.3 billion, so this was a quite respectable study. Nevertheless, for 83 male cells and 163 female cells, there were exposures but no deaths. In most cases these occurred at younger issue ages (below 15) or at higher ages (above 70) where little business is sold and thus few deaths are expected. In the suggested weighting, this would give these cells infinite weighting, which is a practical problem.

Beyond this immediate practical problem, I am concerned about any cell in which few deaths are experienced. Such cells are notorious for outlying values. It seems to me that in such cases, the experience mortality rate may be a biased parameter to use in estimating the variance of such a cell. If the deviation is to an unusually high amount of deaths, that result will be given a low weighting. If the deviation is to a very low amount of deaths, that result will be given a very high weighting. Taken together, that could produce a graduated table with a tendency to systematically underestimate the total deaths.

I have a suggestion that might alleviate these problems, although at the cost of doing twice as many calculations. I suggest that the parametric fit be done in two passes. For the first pass, use the exposures as the weights and calculate a set of preliminary smoothed q factors. Then use these preliminary q factors as the divisor of the exposures for the second pass.

I also have a general question about any graduation process for select and ultimate tables: What statistical tests should be applied to determine whether the graduated table gives reasonable fit and smoothness?

The U.K. actuarial profession has developed several tests in the Continuous Mortality Investigation Reports (CMIR) work. Perhaps the best example is the paper "On Graduation by Mathematical Formula" by Forfar, McCutcheon and Wilkie [1] to explain the methods used to develop the graduated mortality tables in CMIR 9. Section 9 of the paper covers "Tests of a Graduation" and lists signs test (9.3), runs test (9.4), Kolmogorov-Smirnov test (9.5), serial correlation test (9.6), and the chi-squared test (9.7), along with an overall assessment of the tests (9.8). But the tests listed are for one-dimensional graduation and for studies based on number of lives; tests suitable for two-dimensional graduations based on amounts of insurance are more difficult to define. I hope that some of the talented theoreticians who have contributed so much of value to the *Transactions* will take up this question.

REFERENCES

1. 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.
2. KNORR, F.E. "Multidimensional Whittaker-Henderson Graduation," *TSA* XXXVI (1984): 213–55.

PERRY WISEBLATT:

Dr. Carriere should be commended for his research of parametric models. It is clear that a parametric model that fits the underlying crude data has several advantages over a graduated table.

It should be emphasized that many parametric models were tested, such as the 18-parameter model described in the paper. The 11-parameter model used was chosen because the author determined that it offered the best combination of fit and simplicity while satisfying the monotonicity constraint over a broad range of ages. The model is not necessarily representative of mortality in general. Similar methods applied to other data sets may yield different models. It is possible that for certain data sets or for certain purposes, no parametric formula tested will provide an acceptable approximation to the underlying data.

On another note, it may not be appropriate to use a parametric model to extrapolate rates beyond the range of the crude data. Had crude data been available for issue ages over 70, it is likely that the parameter estimates would be different; perhaps even a different model would have been selected. In addition, there is no general agreement on what underwriting criteria should be used to classify older lives as standard risks. The level of mortality measured at the older ages will certainly reflect those judgments.

The breakdown of the monotonicity constraint at the extreme older ages as documented by the author is an indication that, at the very least, caution should be exercised when extrapolating rates beyond the limits of the crude data.

(AUTHOR'S REVIEW OF DISCUSSIONS)

JACQUES F. CARRIERE:

I thank Messrs. Evans, Lumsden and Wiseblatt for their discussions. Mr. Evans points out the lack of fit of my model at various ages, while Mr. Lumsden makes several comments on the appropriateness of the weighting factors used in the loss function. Lastly, Mr. Wiseblatt cautions readers about using the model to extrapolate mortality rates beyond the range of the crude data. Let me respond to each discussant's remarks.

In Figure 5, I think it is obvious that the model systematically underestimates the male rates between the ages of 36 and 53 and overestimates the female rates between the ages of 79 and 99. This lack of fit is the penalty that we must pay for using a parametric model that yields smooth rates. The objective of my paper is not to develop a graduation technique that fits the data everywhere. Instead, I present a parametric model that fits the "overall" pattern of mortality, thereby enabling practitioners to predict the select rates at issue ages above 70 and beyond the 15-year select period. Mr. Evans states that these problems with fit would be "excessive in practice." I claim that using a parametric model is more practical than using tabular rates.

I agree with Mr. Lumsden's comments and suggestions for setting weights. There is no "right" method for choosing the weights, but the "double-pass" technique that Mr. Lumsden presents is a great idea. Essentially, the key to setting good weights is knowing the variance associated with any crude rate. Therefore, I suggest that in the future, the reports prepared by the Society of Actuaries give the variances associated with all the crude rates.

In conclusion, I must agree with Mr. Wiseblatt's comments. Certainly, the parametric models presented here may not fit all select and ultimate data sets, but they do present a starting point for other researchers who may find better models. Notwithstanding the cautions about the predictive ability of this model, my parametric formula is currently the only tool available for predicting the select rates at issue ages above 70 and beyond the 15-year select period.