TRANSACTIONS OF SOCIETY OF ACTUARIES
1978 VOL. 30

A LINEAR PROGRAMMING APPROACH TO GRADUATION*

DONALD R. SCHUETTE

ABSTRACT

The Whittaker-Henderson Type B method of graduation, in which the weighted sum of the squares of the deviations of graduated values from observed values plus a parameter times the sum of the squares of the zth differences of the graduated values is minimized, is modified by using absolute values instead of squares. The resulting problem is expressed as a linear programming problem, and the nature of the solution for various values of the parameter is discussed. Two theorems concerning the optimality of the perfect-fit and perfect-smoothness graduations are proved. Two examples are then presented, and some difficulties are seen to exist in regard to computational feasibility.

INTRODUCTION

All students of actuarial graduation eventually become familiar with the difference-equation method of graduation [9]. In that method, graduated values u_x, $x = 1, 2, \ldots, n$, are sought corresponding to a given set of observed or ungraduated values u_x'' and nonnegative weights w_x that minimize the quantity $F + \theta S$, where $F = \Sigma_{x=1}^{n} w_x(u_x'' - u_x)^2$ and $S = \Sigma_{x=1}^{n-z} (\Delta^z u_x)^2$. F is an expression that measures the degree of fit (actually, lack of fit) of the graduated values to the observed values, and S is an expression that measures the degree of smoothness (lack of smoothness) exhibited by the graduated values. The order of differences used in the measure of smoothness is denoted by z. The values of z commonly employed are $z = 2$ and $z = 3$. The parameter θ is a nonnegative constant that indicates the weighting assigned to smoothness as a desirable characteristic of the graduated values relative to how well they fit the observed values. The larger the value of θ, the smaller S will be and the smoother will be the graduated values.

* This research was performed in part while the author was a visiting member of the Mathematics Research Center, University of Wisconsin, Madison, and was sponsored by the United States Army under Contract No. DAAG 29-75-C-0024.

The author gratefully acknowledges the valuable advice and counsel of Professors T. N. E. Greville and James C. Hickman in the course of preparing this paper.

The method is called the difference-equation method because the values u_x for which the minimum of $F + \theta S$ is achieved can be shown to satisfy the difference equation

$$w_x u_x + \theta\delta^{2z} u_x = w_x u_x'' ,$$

where δ denotes the difference taken centrally. Whittaker [15] first proposed the method, and Henderson [6] developed practical procedures for obtaining the graduated values based on a factorization of the difference equation. The formulas that result in the case where $w_x = 1$ for all x are called Whittaker-Henderson Type A graduation formulas, and those in the more general case of variable weights are called Whittaker-Henderson Type B formulas. Greville [5] has developed very elegantly the general case employing matrix and vector notation, making use of results known in the field of linear algebra.

Basic to the difference-equation method are the choices in the objective function, $F + \theta S$, of the measures of fit and smoothness. Used as the measure of fit is the weighted sum of the *squares* of the deviations, $u_x'' - u_x$, of the observed values from the graduated values and, as the measure of smoothness, the sum of the *squares* of the zth differences of the graduated values.

Other choices are available, however. The objective is that in some sense the deviations and the zth differences be small. A direct and obvious measure is the unsigned magnitude, or absolute value, of those quantities. In his monograph, Miller [9] included the sum of the absolute values of the third differences of the graduated values as one of the "usual" measures of the smoothness of a graduation. Thus, there is precedent for considering absolute values rather than squares. The fact remains, however, that graduation methods based on minimizing sums of squares have been developed extensively, while those based on minimizing sums of absolute values have not.

A first and perhaps minor reason why squares have received more attention than absolute values is that the absolute-value function presents an algebraic sign difficulty similar to the one encountered in statistics in connection with the problem of defining a measure of dispersion in a frequency distribution. Using squares provides one simple solution to that difficulty, a solution that leads to the variance–standard deviation measure of dispersion in the case of the statistical problem and to Whittaker-Henderson methods in the case of graduation.

Fortunately, a method of coping directly with the absolute-value function is available in linear programming. For example, the problem of fitting a line, $u_x = a + bx$, to a given set of observed values so as to minimize the expression

$$S(a, b) = \sum_{x=1}^{n} w_x | u_x'' - (a + bx) |$$

can be formulated as a linear programming problem (see, e.g., Wagner [13], Barrodale and Roberts [2], Armstrong and Frome [1], and Schuette [11]). In fact, the problem of finding the best-fitting linear combination of any given set of functions so as to minimize the weighted sum of the deviations in absolute value can be formulated as a linear programming problem [2], and polynomials of any degree can be so fitted.

A second and undoubtedly more important reason why methods based on minimizing sums of squares have been favored in graduation is the preeminence of the principle of least squares in statistical theory, which in turn can be traced to the normal distribution. In regression theory, when the errors of observation about the regression curve are assumed to be normally distributed, application of the principle of maximum likelihood leads directly to least squares as the criterion to follow in estimating the parameters of the regression curve. In fact, Whittaker and Robinson [16] gave a Bayesian rationale for the Whittaker procedure, in which the minimization of the sums of squares follows from their assumptions that (1) the true underlying values have a multivariate normal prior distribution and (2) the observed values are independently and normally distributed about the true underlying values.

In recent years, however, statisticians have become concerned that least-squares methods may not give the best results in cases where errors of observation follow distributions that tend to generate "outliers" more frequently than the normal distribution. That concern has led to studies of "robust" estimation methods [10], in which the problem is to develop alternatives to least squares that are less sensitive to outliers. The use of least absolute deviations is one of the alternatives that appears to give better results than least squares when the error distribution has heavier tails than the normal distribution [7].

In view of the misgivings that statisticians are having with respect to least squares and the fact that they are examining least absolute values as an estimation procedure, it is appropriate that actuaries do the same with respect to graduation methods. Hence, this paper will be devoted to the task of adapting linear programming to the graduation problem so that absolute values may be employed in place of squares.

At the outset, it should be noted that the following are involved: the general notion of the distance between two vectors, the distance between the vector of observed values and the vector of graduated values in the case of fit, and the distance between the vector of zth differences of the graduated values and the null vector in the case of smoothness. The concept of distance between two vectors or elements of a linear space has been generalized by mathematicians to include an infinity of possible measures. A closely related idea is that of the norm of a vector or element in the space. For example, if $X = (x_1, x_2, \ldots, x_n)$ is a vector of real or complex numbers, the l_p norm of that vector, denoted by $\|X\|_p$, is defined as

$$\|X\|_p = (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{1/p}$$

for $p \geq 1$ (see Davis [4]). If $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$ are two n-component vectors, then the p-distance between them is the l_p norm of their difference, that is, $\|X - Y\|_p$. The l_2 norm of $X - Y$ is

$$\|X - Y\|_2 = \sqrt{\left[\sum_{i=1}^{n} (x_i - y_i)^2\right]},$$

which is the familiar Euclidean distance between the two points in n-dimensional space. The l_1 norm of $X - Y$ is

$$\|X - Y\|_1 = \sum_{i=1}^{n} |x_i - y_i|,$$

which may be described as the rectangular distance between X and Y, or the distance between them when movement is constrained to be always parallel to one of the coordinate axes.

In the Whittaker-Henderson Type A graduation problem with the objective function

$$F + \theta S = \sum_{x=1}^{n} (u''_x - u_x)^2 + \theta \sum_{x=1}^{n-z} (\Delta^z u_x)^2,$$

the terms F and S are the squares of the l_2 norms of the vectors $u'' - u = (u''_1 - u_1, u''_2 - u_2, \ldots, u''_n - u_n)$ and $\Delta^z u = (\Delta^z u_1, \ldots, \Delta^z u_{n-z})$, respectively. In the corresponding objective function to be considered in this paper,

$$F + \theta S = \sum_{x=1}^{n} |u''_x - u_x| + \theta \sum_{x=1}^{n-z} |\Delta^z u_x|,$$

the terms F and S are the l_1 norms of those vectors. When the weights w_x are added to the problem, the resulting expressions for F are not as clearly identifiable as norms of the vector $u'' - u$, but for the purposes of this paper it will be convenient to refer to the traditional Whittaker-Henderson development as being in the l_2 norm and to refer to the development in this paper as being in the l_1 norm.

LINEAR PROGRAMMING FORMULATION OF THE l_1 NORM GRADUATION PROBLEM

The key item in the formulation of the l_1 norm graduation problem is the device for coping with absolute values. That device is the separation of any function into its positive and negative parts. For example, for the function $f(x)$, let

$$P(x) = f(x) \quad \text{if } f(x) \geq 0$$
$$= 0 \quad \text{if } f(x) < 0 \,,$$

and let

$$N(x) = -f(x) \quad \text{if } f(x) < 0$$
$$= 0 \quad \text{if } f(x) \geq 0 \,.$$

Then $f(x)$ always may be replaced by $P(x) - N(x)$, and $|f(x)|$ by $P(x) + N(x)$. The components $P(x)$ and $N(x)$ each must be constrained to be nonnegative, and they must not both be positive simultaneously if the structure of the problem is to be reflected properly. Fortunately, the simplex algorithm of linear programming is ideally suited to comply with those requirements.

For the graduation problem, let

$$u_x'' - u_x = P_x - N_x \,, \tag{1}$$

with $P_x \geq 0$ and $N_x \geq 0$ for $x = 1, 2, \ldots, n$, and let

$$\Delta^z u_x = R_x - T_x \,, \tag{2}$$

with $R_x \geq 0$ and $T_x \geq 0$ for $x = 1, 2, \ldots, n - z$. Then

$$u_x = u_x'' - (P_x - N_x) \tag{3}$$

and

$$\Delta^z u_x = \Delta^z(u_x'' - P_x + N_x) = R_x - T_x \,. \tag{4}$$

Also, $|u_x'' - u_x| = P_x + N_x$ and $|\Delta^z u_x| = R_x + T_x$.

The complete problem then is to find values of P_x, N_x, R_x, and T_x so as to minimize

$$F + \theta S = \sum_{x=1}^{n} w_x(P_x + N_x) + \theta \sum_{x=1}^{n-z} (R_x + T_x) \tag{5}$$

subject to the constraints

$$\Delta^z(P_x - N_x) + R_x - T_x = \Delta^z u_x'' \,, \quad x = 1, 2, \ldots, n - z \,, \tag{6}$$

and $P_x \geq 0$, $N_x \geq 0$, $R_x \geq 0$, and $T_x \geq 0$ for all appropriate values of x.

It should be noted that the constraint equation (6) is obtained by rearranging equation (4) so as to place all the unknowns or decision variables on the left-hand side and the known or directly computable quantities on the right-hand side. Because the operator Δ^z is linear and the variables appear linearly in all terms in equations (5) and (6), the problem is a linear programming problem. The problem involves $2n + 2(n - z)$ variables and $n - z$ constraints. For values of n large enough to span the range of relevant ages in a mortality table, say $n \geq 80$, the problem that results is quite large.

It will be convenient to express the problem in matrix and vector notation similar to that employed by Greville [5]. Let u'', u, P, N, R, and T denote column vectors with components u''_x, u_x, P_x, N_x, R_x, and T_x, respectively. Also, let W denote the column vector of weights w_x. The differencing matrix K_z is the $(n - z) \times n$ matrix such that $K_z u$ is the column vector with components $\Delta^z u_x$ for $x = 1, 2, \ldots, n - z$. For example,

$$K_2 = \begin{bmatrix} 1 & -2 & 1 & 0 \ldots\ldots & 0 \\ 0 & 1 & -2 & 1 & 0 \ldots 0 \\ 0 & 0 & 1 & -2 & 1 \ldots 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 \ldots 1 & -2 & 1 \end{bmatrix}.$$

The problem then may be written as follows:

$$\text{Minimize} \qquad F + \theta S = W^t(P + N) + \underline{\theta}^t(R + T) \tag{7}$$

$$\text{subject to} \qquad K_z(P - N) + I_{n-z}(R - T) = K_z u'' , \tag{8}$$

where $\underline{\theta}$ is a column vector of $n - z$ components all equal to θ, the superscript t denotes the transpose of a vector or matrix, and I_{n-z} is the identity matrix of order $n - z$. All variables are constrained to be nonnegative.

The constraint equations and the matrix of coefficients by means of which those equations are represented play important roles in linear programming. The system of equations (8) may be expressed in the form $AX = K_z u''$, where A is the $(n - z) \times (4n - 2z)$ matrix of coefficients that in partitioned form is

$$A = \left[K_z \;\vdots\; -K_z \;\vdots\; I_{n-z} \;\vdots\; -I_{n-z} \right] \tag{9}$$

and

$$X = \begin{bmatrix} P \\ N \\ R \\ T \end{bmatrix}$$

is the vector of all $4n - 2z$ variables that appear in the problem. Each column of A is associated with one of the variables of the vector X, in that it contains the coefficients of that variable in the system of equations (8). The blocks K_z, $-K_z$, I_{n-z}, and $-I_{n-z}$ are associated respectively with the variables contained in the vectors P, N, R, and T.

LINEAR PROGRAMMING THEORY

With the graduation problem in the l_1 norm expressed as a linear programming problem, the theory and results known in that field can be exploited in the examination of the problem and its solutions. Because some readers may not be familiar with linear programming, this section contains certain definitions and a review of some of the principal theorems.

Linear programming, a special case of the more general field of mathematical programming or constrained optimization, has been developed extensively since its first appearance during World War II. Many excellent treatments of it are available; see, for example, Dantzig [3], Hillier and Lieberman [8], Simonnard [12], and Wagner [14]. The following is taken mainly from Simonnard and pertains to the standard form problem

$$\text{Minimize} \qquad Z = \sum_{j=1}^{n} c_j x_j \tag{10}$$

$$\text{subject to} \quad \sum_{j=1}^{n} a_{ij} x_j = b_i\,, \qquad i = 1, 2, \ldots, m\,, \tag{11}$$

$$\text{and} \qquad x_j \geq 0\,, \quad j = 1, 2, \ldots, n\,. \tag{12}$$

In matrix and vector form the same problem is written

$$\text{Minimize} \qquad Z = C'X \tag{13}$$

$$\text{subject to} \quad AX = b\,, \tag{14}$$

$$X \geq \underline{0}\,, \tag{15}$$

where X is the n-component column vector of unknowns or decision variables; C is the corresponding column vector of objective function coefficients, so that C' is a row vector; Z is the objective function to be minimized; A is the $m \times n$ matrix of constraint-equation coefficients; and b is the m-component column vector of constants that appear on the right-hand side of the constraint equations. $X \geq \underline{0}$ is a compact way of writing $x_j \geq 0$ for $j = 1, 2, \ldots, n$. The symbol $\underline{0}$ represents a column vector with all components equal to zero. It is assumed that $m < n$ and that the rank of A is m, which means that the system of equations (14) is nonredundant.

A basis B is a submatrix of A consisting of m linearly independent columns of A. The variables associated with the columns of B are called the ***basis variables*** and constitute a subvector X_B of X. The corresponding subvector of elements of C is denoted by C_B. When the remaining or nonbasic variables, X_R, are arbitrarily assigned zero values, the system of equations (14) reduces to $BX_B = b$, which has the unique solution

$$X_B = B^{-1}b \, . \tag{16}$$

The solution $X_B = B^{-1}b$ and $X_R = \underline{0}$ is called a *basic solution*, and the value of the objective function Z corresponding to this basic solution is $Z = C_B^t X_B$, which from equation (16) becomes $Z = C_B^t B^{-1}b$.

The values of the basic variables given by equation (16) may or may not be all nonnegative for an arbitrarily selected basis; if they are, the solution is called a *basic feasible solution*. If the value of Z corresponding to a basic feasible solution is less than or equal to the values of Z for all other feasible solutions, the solution is an *optimal solution*.

What Simonnard calls the fundamental theorem of linear programming may be stated as follows:

> *For the standard form linear programming problem,* (a) *if it has at least one finite feasible solution, then it has at least one basic feasible solution;* (b) *if it has at least one finite feasible solution that is optimal, then it has at least one basic feasible solution that is optimal.*

The importance of the fundamental theorem is that it justifies procedures that confine the search for optimal solutions to the set of basic feasible solutions. The simplex algorithm is one such procedure. It progresses iteratively from one basic feasible solution to another that is better until no further improvement in the value of the objective function can be achieved. An optimal basic feasible solution then has been found.

Because the feasible region or set of all feasible solutions is the intersection of linear half-spaces, it is a convex set. Moreover, when the feasible region is bounded, the following theorem applies:

> *If the feasible region for the standard linear programming problem is bounded, there is at least one basic feasible solution that is optimal.*

LINEAR PROGRAMMING THEORY APPLIED TO THE GRADUATION PROBLEM

For the l_1 norm graduation problem given by equations (8) and (9), a number of conclusions may be drawn immediately in light of the preceding discussion. First of all, a finite feasible solution exists, namely, $u_x = u_x''$ for all x, which is the solution in which there is no graduation at all. In that case, $F = 0$ and $F + \theta S = \theta \Sigma_{x=1}^{n-z} |\Delta^z u_x''|$. Furthermore, the no-graduation solution is a basic feasible solution, the basic variables being R_x if $\Delta^z u_x'' \geq 0$ and T_x if $\Delta^z u_x'' < 0$, $x = 1, 2, \ldots, n - z$. Then $P_x = N_x = 0$ for all x; $R_x = \Delta^z u_x''$ if $\Delta^z u_x'' \geq 0$, and $R_x = 0$ if $\Delta^z u_x'' < 0$. $T_x = -\Delta^z u_x''$ if $\Delta^z u_x'' < 0$, and $T_x = 0$ if $\Delta^z u_x'' \geq 0$.

Thus, the graduation problem has at least one basic feasible solution. Moreover, the existence of that solution leads to the conclusion that an optimal basic feasible solution exists, because, although the feasible region for the original problem is unbounded, that problem can be revised by the addition of the constraint

$$\sum_{x=1}^{n} w_x(P_x + N_x) + \theta \sum_{x=1}^{n-z} (R_x + T_x) \leq \theta \sum_{x=1}^{n-z} |\Delta^z u_x''| .$$

Since all the coefficients w_x and θ are nonnegative, the effect of the added constraint is to bound each of the variables and, hence, to bound the feasible region. In view of the theorem stated at the end of the preceding section, it may be concluded that the revised problem has an optimal basic feasible solution, from which it follows that the original problem has one as well.

A second conclusion is that, in any basic solution, $u_x = u_x''$ for at least z values of x. This follows from the fact that in any basic solution at most $n - z$ of the variables P_x and N_x are basic and can have positive values. Hence, for at least z values of x, both P_x and N_x are nonbasic and have value zero, in which case $u_x = u_x''$ because, in general, $u_x'' - u_x = P_x - N_x$. It also should be observed that, for any value of x, P_x and N_x will not both be positive in any basic solution, because the associated columns in A are the negatives of each other and cannot both be contained in a set of linearly independent columns forming a basis.

TWO CRITICAL VALUES OF θ

Miller [9] has observed that for the graduation problem in the l_2 norm the optimal solution when $\theta = 0$ is the no-graduation solution, and that as θ grows very large the optimal values of u_x approach values lying on the line fitted to the values of u_x'' by the method of weighted least squares when $z = 2$. The result for general values of z is that as θ grows large the graduated values approach those given by a polynomial in x of degree $z - 1$ fitted by the method of weighted least squares.

In the case of the l_1 norm the situation, although similar, is different in a way that might be said to characterize the l_1 norm and its linear programming formulation. In this case it will be shown that there are two critical values of θ, θ_L and θ_U, such that for $0 \leq \theta \leq \theta_L$ the no-graduation solution is optimal, while for $\theta \geq \theta_U$ the optimal values of u_x are those lying on the polynomial of degree $z - 1$ fitted to the observed values by the method of least total absolute deviations. The determination of those critical values, θ_L and θ_U, is one of the principal objectives of this paper. Formulas for θ_L and θ_U are embodied in the two theorems that will be proved. The proofs, however, require knowledge of some of the details of the simplex algorithm of linear programming, which will be discussed now.

A critical element in the simplex algorithm is the notion of the reduced cost or evaluator of a nonbasic variable in a basic feasible solution. The reduced cost of a nonbasic variable may be described as the reduction in the objective function per unit increase in the variable that currently has value zero. (Some authors prefer to deal with the negative of the quantity described here, that is, the increase in the objective function per unit increase in the variable.) By examination of the reduced costs of all nonbasic variables it can be determined whether or not further improvement in the objective function can be achieved. If all reduced costs for nonbasic variables are zero or negative, no improvement can be made in a minimizing problem and the current solution is optimal. If the reduced cost of at least one nonbasic variable is positive, that variable possibly can be made basic and the objective function reduced. The process of deciding which variable to make nonbasic, of determining the new values of the basic variables so as to maintain feasibility, and of recomputing the revised reduced costs for the nonbasic variables constitutes one iteration of the simplex algorithm.

The task of computing the reduced costs for the standard problem of equations (10)–(12) now will be examined. Consider the constraint matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

Suppose that a linear combination of the rows is obtained by multiplying the first row by y_1, the second row by y_2, and so on, and adding. Let $z_j = \Sigma_{i=1}^{m} y_i a_{ij}$ be the quantity obtained for the jth column. Furthermore, suppose that multipliers y_i have been determined so that for a preselected set of m basic columns the equations $z_j = c_j$ hold. Such a set of multipliers indeed can be found whenever the columns are linearly independent, which they must be to form a basis. Then for a nonbasic column the quantity $z_j - c_j$ is the reduced cost for the associated nonbasic variable x_j. Notice that for the basic variables $z_j - c_j = 0$.

The argument that $z_j - c_j$ can be interpreted as the decrease in the objective function per unit change in the variable x_j proceeds as follows. Let B denote the matrix of basic columns of A. Then, as seen previously, $X_B = B^{-1}b$ is the vector of values of the basic variables and $Z = C_B^t X_B$ is the value of the objective function. Let A_j denote the column of A corresponding to x_j, and let $A_j = BT_j$, which means that T_j is the vector of coefficients in the expression for A_j as a linear combination of the columns of B. If x_j becomes positive in value, the values of the currently basic variables have to be modified in order that the equation $AX = b$ continue to hold. The previous equation was $BX_B = b$. Since $A_j = BT_j$, the addition of the equation $x_j(A_j - BT) = 0$ to the previous equation produces the result $B(X_B - x_jB^{-1}A_j) + x_jA_j = b$. This result may be interpreted to mean that $X_B' = X_B - x_jB^{-1}A_j$ is the vector of modified values of the basic variables. The modified value of the objective function is

$$Z' = C_B^t X_B' + c_j x_j .$$

Substitution for X_B' produces

$$\begin{aligned} Z' &= C_B^t(X_B - x_jB^{-1}A_j) + c_jx \\ &= Z - x_j(C_B^tB^{-1}A_j - c_j) \\ &= Z - x_j(z_j - c_j) \end{aligned}$$

because the vector of multipliers $Y_B^t = (y_1, y_2, \ldots, y_m)$ must satisfy the equation $Y_B^tB = C_B^t$ or $Y_B^t = C_B^tB^{-1}$, and therefore $z_j = C_B^tB^{-1}A_j$. Hence, $z_j - c_j$ represents the change in the objective function value per unit change in the value of the corresponding variable x_j.

Readers who are familiar with the dual problem in linear programming will recognize the multipliers y_i as the dual variables. It should be mentioned that some of the later sections of this paper could have been presented in terms of the dual problem and its variables, but the author has chosen not to do so.

The first of the two theorems follows:

THEOREM 1. *If $0 \le \theta \le \theta_L$, then the values of u_x, $x = 1, 2, \ldots, n$, for which the quantity*

$$F + \theta S = \sum_{x=1}^{n} w_x|u_x'' - u_x| + \theta \sum_{x=1}^{n-z} |\Delta^z u_x|$$

is a minimum are $u_x = u_x''$, $x = 1, 2, \ldots, n$, *where*

$$\theta_L = \underset{x=1,2,\ldots,n}{\text{minimum}} \{w_x / |\Delta^z v_{x-2}|\}$$

and the quantities $\Delta^z v_x$ *are the zth differences of the sequence* v_x *defined as follows:*

For $x = 1, 2, \ldots, n - z$,

$$v_x = 1 \qquad \text{if } \Delta^z u_x'' \geq 0$$
$$= -1 \qquad \text{if } \Delta^z u_x'' < 0 \,.$$

For $x = -z + 1, -z + 2, \ldots, 0$ *and* $x = n - z + 1, \ldots, n$,

$$v_x = 0 \,.$$

Before the proof of Theorem 1 is presented, the procedure for computing θ_L that is implied by the theorem will be illustrated in connection with the example employed by Greville [5, p. 59].

EXAMPLE: CALCULATION OF θ_L FOR $z = 2$

x	u_x''	w_x	$\Delta u_x''$	$\Delta^2 u_x''$	v_x	Δv_{x-1}	$\Delta^2 v_{x-2}$	$w_x / \|\Delta^2 v_{x-2}\|$
−1	......	......	......	......	0	......	......	......
0	......	......	......	......	0	0	......	......
1	34	3	−10	17	1	1	1	3
2	24	5	7	2	1	0	−1	5
3	31	8	9	−19	−1	−2	−2	4
4	40	10	−10	29	1	2	4	2.5
5	30	15	19	−20	−1	−2	−4	3.75
6	49	20	−1	1	1	2	4	5
7	48	23	0	19	1	0	−2	11.5
8	48	20	19	−28	−1	−2	−2	10
9	67	15	−9	18	1	2	4	3.75
10	58	13	9	......	0	−1	−3	4.33
11	67	11	......	......	0	0	1	11

The values of v_x are merely the signs of $\Delta^z u_x''$ with z additional zero values appended at the beginning and at the end. In this example the minimum value of $w_x/|\Delta^z v_{x-2}|$ occurs for $x = 4$. Thus $\theta_L = 2.5$.

Proof of Theorem 1: Consider the constraint matrix for the problem,

$$\begin{array}{ccccc} & P^t & N^t & R^t & T^t \\ A = & \left[K_z \right. & -K_z & I_{n-z} & \left. -I_{n-z} \right], \\ & W^t & W^t & \underline{0}^t & \underline{0}^t \end{array}$$

where the variables associated with each block of A are indicated as a row vector above that block and the corresponding objective function coefficients as a row vector below that block. Let $y_1, y_2, \ldots, y_{n-z}$ be multipliers for the respective rows. Because the blocks with which the variables R_j and T_j are associated are the identity matrix and its negative, the reduced costs for those variables are

$$Z_{R_j} - C_{R_j} = y_j - \theta \quad \text{and} \quad Z_{T_j} - C_{T_j} = -y_j - \theta \,.$$

It should be noted that for any set of multipliers the sum of the reduced costs for R_j and T_j equals -2θ. Suppose that the values of the multipliers are determined so that the reduced costs of the basic variables corresponding to the no-graduation solution equal zero, that is,

$$y_j = \theta \qquad \text{if } \Delta^z u_j'' \geq 0 \text{ and } R_j \text{ is basic} \tag{17}$$

$$= -\theta \qquad \text{if } \Delta^z u_j'' < 0 \text{ and } T_j \text{ is basic}\,. \tag{18}$$

The reduced costs for all of the variables R_j and T_j then have values equal to either zero or -2θ.

Consider next the reduced costs for the variables P_j and N_j. Except for $j = 1, 2, \ldots, z$ and $j = n - z + 1, n - z + 2, \ldots, n$, the reduced costs have the form

$$Z_{P_j} - C_{P_j} = (-1)^z \Delta^z y_{j-z} - w_j \tag{19}$$

and

$$Z_{N_j} - C_{N_j} = (-1)^{z+1} \Delta^z y_{j-z} - w_j \,. \tag{20}$$

However, if values $y_j = 0$ are assigned for $j = -z + 1, -z + 2, \ldots, 0$ and $j = n - z + 1, n - z + 2, \ldots, n$, equations (19) and (20) hold for $j = 1, 2, \ldots, n$. However, in view of equations (17) and (18),

$$Z_{P_j} - C_{P_j} = \theta(-1)^z \Delta^z v_{j-z} - w_j \tag{21}$$

and

$$Z_{N_j} - C_{N_j} = \theta(-1)^{z+1} \Delta^z v_{j-z} - w_j \,, \tag{22}$$

where the values v_j are defined as indicated in the statement of the theorem. The conditions for this basic solution to be optimal are $Z_{P_j} - C_{P_j} \leq 0$ and $Z_{N_j} - C_{N_j} \leq 0$ for $j = 1, 2, \ldots, n$. In view of equations (21) and (22), the latter conditions are equivalent to

$$-w_j < \theta(-1)^z \Delta^z v_{j-z} < w_j \quad \text{or} \quad |\theta \Delta^z v_{j-z}| < w_j \,. \tag{23}$$

But inequality (23) will be satisfied if $\theta \leq \theta_L$, where

$$\theta_L = \underset{j=1,2,\ldots,n}{\text{minimum}} \{w_j / |\Delta^z v_{j-z}|\} \,. \tag{24}$$

Hence, if $0 \leq \theta \leq \theta_L$, the no-graduation solution $u_x = u_x''$ is the optimal solution.

The second of the two theorems will be approached from what will be called a polynomial basis, that is, a basis for which the associated basic variables are all selected from the variables P_j and N_j, $j = 1, 2, \ldots, n$. Such a basis is called a polynomial basis because none of the variables R_j and T_j are basic; therefore, their values all equal zero, all zth differences of the graduated values equal zero, and the graduated values lie on a polynomial of degree $z - 1$ or lower. Since there are only $n - z$ basic variables, there must be at least z values of j such that neither P_j nor N_j is basic. For those values of j, $u_j = u_j''$, and the polynomial of degree $z - 1$ passes through those points.

Suppose that, for a polynomial basis, multipliers $y_1, y_2, \ldots, y_{n-z}$ are determined so that the reduced costs for the basic variables equal zero, that is,

$$(-1)^z \Delta^z y_{j-z} = w_j \quad \text{if } P_j \text{ is basic} \tag{25}$$

and

$$(-1)^{z+1} \Delta^z y_{j-z} = w_j \quad \text{if } N_j \text{ is basic}\,, \tag{26}$$

with $y_j = 0$ for $j = -z+1, -z+2, \ldots, 0$ and $j = n-z+1, n-z+2, \ldots, n$. The reduced cost for N_j is $-2w_j$ if P_j is basic, and the reduced cost for P_j is $-2w_j$ if N_j is basic.

Suppose further that the reduced costs for all other nonbasic P_j and N_j are also nonpositive, that is, $(-1)^z \Delta^z y_{j-z} \leq w_j$ and $(-1)^{z+1} \Delta^z y_{j-z} \leq w_j$, or, in other words,

$$-w_j < \Delta^z y_{j-z} < w_j\,, \quad j = 1, 2, \ldots, n\,. \tag{27}$$

The basic solution so determined represents the solution to the problem of fitting a polynomial of degree $z - 1$ or lower to the observed values u_x'' so as to minimize $\Sigma_{x=1}^{n} w_x | u_x'' - u_x |$. That solution is also the solution to the graduation problem if θ is sufficiently large. The question of how large θ must be is answered by the following theorem.

THEOREM 2. *Let $y_1, y_2, \ldots, y_{n-z}$ be multipliers corresponding to a polynomial basis and satisfying conditions (25), (26), and (27). Let*

$$\theta_U = \underset{j=1,2,\ldots,n-z}{\text{maximum}} \{|y_j|\}\,.$$

Then, if $\theta \geq \theta_U$, the associated basic solution is optimal for the graduation problem.

Proof: By hypothesis the optimality conditions for the variables P_j and N_j, $j = 1, 2, \ldots, n$, are satisfied, so all that remains to be shown is that the reduced costs for the variables R_j and T_j, $j = 1, 2, \ldots, n - z$, are nonpositive. Those reduced costs are $Z_R - C_R = y_j - \theta$ and $Z_{T_j} -$

$C_{T_j} = -y_j - \theta$. The conditions for all to be nonpositive are $-\theta \leq y_j \leq \theta$, or $|y_j| \leq \theta, j = 1, 2, \ldots, n - z$. Those conditions are satisfied if

$$\theta \geq \underset{j=1,2,\ldots,n-z}{\text{maximum}} \{|y_j|\} .$$

LINEAR PROGRAMMING GRADUATION STRATEGY

Striking a balance between fit and smoothness is the central issue in graduation. In the difference-equation method, that problem takes the form of selecting an appropriate value for θ. Heretofore little guidance has been given the graduator as to how the value of θ might be chosen. The two theorems of the preceding section perhaps offer some help in the case of l_1 norm graduation. The value of θ_L can be computed directly from the data that are originally available. The value of θ_U is not obtained quite so readily, however. First it is necessary to solve the problem of fitting a polynomial of degree $z - 1$ or lower to the data by the method of least total deviations, that is, so as to minimize $\Sigma_{x=1}^{n} w_x |u_x'' - u_x|$. As indicated earlier, the latter problem could be formulated as a separate linear programming problem and solved. However, the linear programming formulation of the graduation problem can be employed just as well. One simply selects a very large value for θ and then runs the graduation problem program. For a value of θ sufficiently large, the variables R_j and T_j all will be nonbasic at optimum. From their reduced costs, which most computer linear programming routines generate, the values of y_j can be obtained. Theorem 2 then can be employed to compute θ_U.

One way to proceed after θ_L and θ_U are known is to solve the linear programming problem for a number of values of θ between θ_L and θ_U. It is very likely that the resulting graduations will permit the user to select a suitable set of graduated values. No single criterion can be offered for the selection of one final set of graduated values. However, the methods that have been proposed provide the graduator with guidelines that can be helpful in exploring the range between perfect fit and perfect smoothness, and also provide control over the problem of striking a balance between those extremes.

EXAMPLES OF THE TECHNIQUE

Example I is that given by Miller [9, p. 39], in which there are nineteen values to be graduated. The first eleven of those values form the example that was employed by Greville [5] and for which the value of θ_L was computed earlier. A computer program that first computes θ_L and then solves the linear programming problem expressed by equations (7) and (8) for $\theta = 10^5$ was employed for $z = 2, 3$, and 4. For $z = 2$ the problem contained 17 constraints and 72 variables, for $z = 3$ it contained 16 constraints and 70 variables, and for $z = 4$ there were 15 constraints and 68 variables. In each case the resulting graduated values were those given by a polynomial of degree $z - 1$. The program utilized a linear programming routine[1] that made available the reduced costs of the variables R_j and T_j from which the values of θ_U were obtained.

The next step was to divide the interval from θ_L to θ_U into ten subintervals and to solve the linear programming problem for those intermediate values of θ. The resulting graduated values for $z = 2$, 3, and 4 are shown in Tables 1, 2, and 3, respectively, along with the measures of fit, $\Sigma\, w_x(u_x'' - u_x)^2$ and $\Sigma\, w_x|u_x'' - u_x|$, and the measures of smoothness, $\Sigma\, (\Delta^z u_x)^2$ and $\Sigma\, |\Delta^z u_x|$. The values of θ_L and θ_U are as follows:

z	θ_L	θ_U
2	1.00	79.00
3	0.75	62.36
4	0.50	11.31

The thirty graduations, including the computation of measures of fit and smoothness, required 58.495 seconds of CPU time for compilation, collection, and execution on the UNIVAC 1110 system.

Some observations concerning the graduated values can be made. First, different values of θ produce the same graduations in a number of instances. For $z = 2$ the graduations are the same for $\theta = 24.40$–55.60, inclusive, and also the two graduations for $\theta = 63.40$ and $\theta = 71.20$ are the same. For $z = 3$ the ten values of θ employed result in only four different graduations. For $z = 4$ three of the graduations are identical.

A second point is that the higher the value of z the relatively farther on the interval from θ_L to θ_U the value of θ must be in order to achieve an acceptable graduation. For $z = 2$, only the second graduation, corresponding to $\theta = 16.60$, appears acceptable. For $z = 3$ each of the two intermediate graduations, corresponding to $\theta = 13.07$–56.20, appears reasonable. For $z = 4$ the last three graduations, corresponding to $\theta =$ 9.15, 10.23, and 11.31, appear to be good. The same point also can be made in either of the following ways: (1) the lower the value of z, the more quickly do the graduated values react to the smoothness requirement and tend toward values lying on a polynomial of degree $z - 1$ as θ increases over the interval from θ_L to θ_U, and (2) the higher the value of z, the better fitting is the graduation corresponding to $\theta = \theta_U$ and the

[1] SIMPLX, a FORTRAN callable subroutine available on the UNIVAC 1110 system of the University of Wisconsin—Madison.

TABLE 1

EXAMPLE I: GRADUATED VALUES, AND MEASURES OF FIT AND SMOOTHNESS; $z = 2$

x	u_x''	w_x	$\theta=8.80$	$\theta=16.60$	$\theta=24.40$	$\theta=32.20$	$\theta=40.00$	$\theta=47.80$	$\theta=55.60$	$\theta=63.40$	$\theta=71.20$	$\theta=79.00$
			Graduated Value (u_x)									
1......	34	3	19.67	22.50	22.50	22.50	22.50	22.50	22.50	21.33	21.33	19.20
2......	24	5	25.33	26.75	26.75	26.75	26.75	26.75	26.75	25.78	25.78	24.00
3......	31	8	31.00	31.00	31.00	31.00	31.00	31.00	31.00	30.22	30.22	28.80
4......	40	10	36.67	35.25	35.25	35.25	35.25	35.25	35.25	34.67	34.67	33.60
5......	30	15	42.33	39.50	39.50	39.50	39.50	39.50	39.50	39.11	39.11	38.40
6......	49	20	48.00	43.75	43.75	43.75	43.75	43.75	43.75	43.56	43.56	43.20
7......	48	23	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00
8......	48	20	48.00	52.25	52.25	52.25	52.25	52.25	52.25	52.44	52.44	52.80
9......	67	15	54.17	57.17	56.79	56.79	56.79	56.79	56.79	56.89	56.89	57.60
10.....	58	13	60.33	62.08	61.33	61.33	61.33	61.33	61.33	61.33	61.33	62.40
11.....	67	11	66.50	67.00	67.00	67.00	67.00	67.00	67.00	67.00	67.00	67.20
12.....	75	10	72.67	71.92	72.67	72.67	72.67	72.67	72.67	72.67	72.67	72.00
13.....	76	9	78.83	76.83	78.33	78.33	78.33	78.33	78.33	78.33	78.33	76.80
14.....	76	9	85.00	84.47	84.00	84.00	84.00	84.00	84.00	84.00	84.00	81.60
15.....	102	7	92.50	92.10	89.67	89.67	89.67	89.67	89.67	89.67	89.67	86.40
16.....	100	5	100.00	99.73	95.33	95.33	95.33	95.33	95.33	95.33	95.33	91.20
17.....	101	5	107.50	107.37	101.00	101.00	101.00	101.00	101.00	101.00	101.00	96.00
18.....	115	3	115.00	115.00	106.67	106.67	106.67	106.67	106.67	106.67	106.67	100.80
19.....	134	1	122.50	122.63	112.33	112.33	112.33	112.33	112.33	112.33	112.33	105.60
			Measures of Fit and Smoothness									
$\Sigma\, w_x(u_x''-u_x)^2$........			7,412.89	6,358.39	7,164.92	7,164.92	7,164.92	7,164.92	7,164.92	7,228.36	7,228.36	8,880.16
$\Sigma\, w_x\lvert u_x''-u_x\rvert$........			756.67	857.20	897.37	897.37	897.37	897.37	897.37	908.56	908.56	1,001.20
$\Sigma\,(\Delta^z u_x)^2$............			71.92	7.82	1.35	1.35	1.35	1.35	1.35	1.49	1.49	0.00
$\Sigma\,\lvert\Delta^z u_x\rvert$............			13.17	3.38	1.42	1.42	1.42	1.42	1.42	1.22	1.22	0.00

TABLE 2

EXAMPLE I: GRADUATED VALUES, AND MEASURES OF FIT AND SMOOTHNESS; $z = 3$

x	u_x''	w_x	$\theta=6.91$	$\theta=13.07$	$\theta=19.23$	$\theta=25.40$	$\theta=31.56$	$\theta=37.72$	$\theta=43.88$	$\theta=50.04$	$\theta=56.20$	$\theta=62.36$
			Graduated Value (u_x)									
1	34	3	20.67	22.32	22.32	22.32	22.32	22.32	23.13	23.13	23.13	30.99
2	24	5	25.99	26.68	26.68	26.68	26.68	26.68	27.01	27.01	27.01	32.87
3	31	8	31.00	31.00	31.00	31.00	31.00	31.00	31.00	31.00	31.00	35.13
4	40	10	35.71	35.29	35.29	35.29	35.29	35.29	35.09	35.09	35.09	37.77
5	30	15	40.11	39.56	39.56	39.56	39.56	39.56	39.29	39.29	39.29	40.80
6	49	20	44.21	43.79	43.79	43.79	43.79	43.79	43.59	43.59	43.59	44.21
7	48	23	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00
8	48	20	52.18	52.18	52.18	52.18	52.18	52.18	52.51	52.51	52.51	52.18
9	67	15	56.73	56.73	56.73	56.73	56.73	56.73	57.13	57.13	57.13	56.73
10	58	13	61.68	61.68	61.68	61.68	61.68	61.68	61.85	61.85	61.85	61.68
11	67	11	67.00	67.00	67.00	67.00	67.00	67.00	67.00	67.00	67.00	67.00
12	75	10	72.71	72.71	72.71	72.71	72.71	72.71	72.58	72.58	72.58	72.71
13	76	9	78.80	78.80	78.80	78.80	78.80	78.80	78.58	78.58	78.58	78.80
14	76	9	85.27	85.27	85.27	85.27	85.27	85.27	85.01	85.01	85.01	85.27
15	102	7	92.13	92.13	92.13	92.13	92.13	92.13	91.87	91.87	91.87	92.13
16	100	5	99.37	99.37	99.37	99.37	99.37	99.37	99.15	99.15	99.15	99.37
17	101	5	106.99	106.99	106.99	106.99	106.99	106.99	106.86	106.86	106.86	106.99
18	115	3	115.00	115.00	115.00	115.00	115.00	115.00	115.00	115.00	115.00	115.00
19	134	1	123.39	123.39	123.39	123.39	123.39	123.39	123.57	123.57	123.57	123.39
			Measures of Fit and Smoothness									
$\Sigma\, w_x(u_x''-u_x)^2$			6,707.94	6,557.30	6,557.30	6,557.30	6,557.30	6,557.30	6,433.84	6,433.84	6,433.84	6,793.25
$\Sigma\, w_x\lvert u_x''-u_x\rvert$			870.01	872.63	872.63	872.63	872.63	872.63	876.08	876.08	876.08	896.17
$\Sigma\, (\Delta^z u_x)^2$			0.47	0.17	0.17	0.17	0.17	0.17	0.10	0.10	0.10	0.00
$\Sigma\, \lvert\Delta^z u_x\rvert$			0.69	0.41	0.41	0.41	0.41	0.41	0.32	0.32	0.32	0.00

TABLE 3

EXAMPLE I: GRADUATED VALUES, AND MEASURES OF FIT AND SMOOTHNESS; $z = 4$

x	u_x''	w_x	$\theta=1.58$	$\theta=2.66$	$\theta=3.74$	$\theta=4.82$	$\theta=5.90$	$\theta=6.99$	$\theta=8.07$	$\theta=9.15$	$\theta=10.23$	$\theta=11.31$
			Graduated Value (u_x)									
1......	34	3	34.00	34.00	34.00	34.00	34.00	34.00	34.00	24.50	24.97	22.22
2......	24	5	24.00	26.47	26.48	26.48	26.48	30.29	30.82	27.53	27.71	26.68
3......	31	8	31.00	31.00	31.00	31.00	31.00	31.00	31.00	31.00	31.00	31.00
4......	40	10	40.00	40.00	40.00	40.00	40.00	34.63	33.67	34.84	34.76	35.24
5......	30	15	46.17	45.89	45.94	45.94	45.94	39.68	37.96	39.00	38.91	39.46
6......	49	20	49.00	48.59	48.66	48.66	48.66	44.63	43.03	43.41	43.35	43.69
7......	48	23	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00
8......	48	20	48.00	48.00	48.00	48.00	48.00	51.27	52.88	52.72	52.78	52.43
9......	67	15	51.50	52.47	52.69	52.69	52.69	55.96	57.67	57.50	57.59	57.04
10.....	58	13	58.00	59.45	59.78	59.78	59.78	61.41	62.38	62.28	62.36	61.88
11.....	67	11	67.00	67.00	67.00	67.00	67.00	67.00	67.00	67.00	67.00	67.00
12.....	75	10	74.37	73.17	72.08	72.08	72.08	72.08	71.54	71.59	71.42	72.45
13.....	76	9	76.00	76.00	76.00	76.00	76.00	76.00	76.00	76.00	76.00	78.28
14.....	76	9	78.54	78.42	79.82	79.82	79.82	79.82	80.89	80.78	81.14	84.55
15.....	102	7	88.66	83.33	84.59	84.59	84.59	84.59	86.73	86.51	87.22	91.30
16.....	100	5	100.00	90.83	91.34	91.34	91.34	91.34	94.02	93.75	94.63	98.60
17.....	101	5	106.22	101.00	101.13	101.13	101.13	101.13	103.27	103.05	103.76	106.48
18.....	115	3	115.00	115.00	115.00	115.00	115.00	115.00	115.00	115.00	115.00	115.00
19.....	134	1	134.00	134.00	134.00	134.00	134.00	134.00	129.72	130.15	128.73	124.22
			Measures of Fit and Smoothness									
$\Sigma\, w_x(u_x''-u_x)^2$.........			8,969.13	9,963.94	9,672.82	9,672.82	9,672.82	7,181.01	6,520.19	6,710.05	6,552.41	6,485.78
$\Sigma\, w_x\lvert u_x''-u_x\rvert$.........			623.63	712.40	725.69	725.69	725.69	822.59	861.87	868.06	869.49	874.85
$\Sigma\, (\Delta^z u_x)^2$............			660.51	138.67	122.66	122.66	122.66	16.45	1.00	0.39	0.22	0.00
$\Sigma\, \lvert\Delta^z u_x\rvert$............			66.61	26.07	21.21	21.21	21.21	6.83	1.37	0.63	0.47	0.00

better fitting are the graduations corresponding to values of θ relatively close to θ_U.

In Example II, the data set to be considered is the 1955–60 ultimate experience from which the 1955–60 Ultimate Basic Mortality Table, Male and Female Lives Combined, was derived. The experience and the table are contained in the Report of the Committee on Mortality under Ordinary Insurances and Annuities appearing in *TSA*, *1962 Reports*. The observed, or crude, mortality rates per 1,000, denoted by u''_x, are given in Table 8 of the report for $x = 15, 16, \ldots, 95$. The actual death claims in units of 1,000 are given in Table 10 of the report. The exposures were not published but can be inferred from the information given, although with considerable loss in the number of significant digits. If $u''_x = q''_x \times 10^3$ denotes the entries in Table 8 and $\theta''_x \times 10^{-3}$ denotes the entries in Table 10, and if w_x is the ratio of the latter to the former, one obtains $w_x = E_x \times 10^{-6}$. Hence, the weights w_x are the exposures in units of 1,000,000.

The first attempt to compute θ_L and θ_U for $z = 2$, 3, and 4 and to obtain graduations at intermediate values of θ ended in failure because the CPU time limit had been exceeded. It was discovered that each call of the linear programming subroutine required approximately three minutes of CPU time on the UNIVAC 1110 system. Hence, the project of obtaining ten graduations for each value of z was abandoned, and instead experimentation with $z = 2$ and $z = 3$ for various values of θ was performed.

For $z = 2$ all the graduations obtained were discarded as being unacceptable for one reason or another. For $z = 3$ several good graduations were obtained. Those graduations, along with the published mortality table from Table 2 of the report, were tested for smoothness and fit. The graduated values and the measures of fit and smoothness are shown in Table 4. The published table was obtained through Whittaker-Henderson Type A methods with $\theta = 3$ $(a = 1)$ for $x \geq 32$ and $\theta = 18$ $(a = 2)$ for $x < 32$. Compared with the published table, the table of graduated values obtained using the methods of this paper for $\theta = 5$ is better with respect to both measures of smoothness and one measure of fit; it is slightly poorer with respect to the measure of fit $\Sigma\, w_x(u''_x - u_x)^2$, which, of course, ought to be small for Whittaker-Henderson graduated values.

CONCLUSIONS

A method of graduation based upon the minimization of the sum of the absolute values of the deviations and the sum of the absolute values of the zth differences of the graduated values has been presented. The

TABLE 4—EXAMPLE II: GRADUATED VALUES, AND MEASURES OF FIT AND SMOOTHNESS; $z = 3$
(1955–60 Ultimate Basic Mortality—Male and Female Lives Combined)

Age	Crude Mortality Rate per 1,000	Exposure in Units of 1,000,000	$\theta=2$	$\theta=3$	$\theta=4$	$\theta=5$	$\theta=10$	$\theta=15$	$\theta=20$	$\theta=35$	Published Mortality Rate per 1,000
			Graduated Value								
15......	0.58	2.69	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.61
16......	0.65	2.91	0.65	0.65	0.67	0.67	0.67	0.66	0.66	0.66	0.68
17......	0.71	2.35	0.73	0.73	0.75	0.75	0.75	0.72	0.73	0.72	0.74
18......	0.83	1.89	0.82	0.81	0.82	0.82	0.82	0.79	0.80	0.79	0.81
19......	0.89	1.62	0.89	0.88	0.88	0.88	0.88	0.84	0.86	0.84	0.86
20......	0.93	1.27	0.94	0.93	0.93	0.93	0.93	0.89	0.91	0.89	0.90
21......	1.07	0.94	0.96	0.96	0.97	0.97	0.97	0.93	0.95	0.93	0.94
22......	0.97	0.89	0.97	0.97	1.00	1.00	1.00	0.97	0.99	0.97	0.95
23......	0.84	0.88	0.95	0.98	1.02	1.02	1.02	1.00	1.01	1.00	0.96
24......	0.90	0.89	0.96	1.00	1.03	1.03	1.04	1.02	1.03	1.02	0.98
25......	0.88	1.85	0.98	1.02	1.04	1.04	1.05	1.04	1.05	1.04	1.00
26......	1.18	2.31	1.02	1.04	1.05	1.05	1.05	1.05	1.05	1.05	1.02
27......	1.05	2.89	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.04
28......	1.31	3.55	1.06	1.06	1.04	1.04	1.04	1.05	1.04	1.05	1.05
29......	0.86	4.26	1.05	1.05	1.03	1.03	1.03	1.05	1.04	1.05	1.05
30......	1.06	5.52	1.06	1.06	1.04	1.04	1.03	1.06	1.06	1.06	1.06
31......	0.99	6.67	1.08	1.07	1.07	1.07	1.06	1.09	1.09	1.09	1.07
32......	1.11	7.82	1.11	1.11	1.11	1.11	1.11	1.13	1.13	1.13	1.10
33......	1.24	9.28	1.14	1.17	1.17	1.17	1.18	1.19	1.19	1.19	1.16
34......	1.14	11.12	1.17	1.25	1.25	1.25	1.26	1.26	1.26	1.26	1.22
35......	1.35	13.19	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.31
36......	1.46	15.40	1.46	1.46	1.45	1.43	1.45	1.45	1.45	1.45	1.42
37......	1.50	17.97	1.50	1.57	1.56	1.50	1.56	1.56	1.56	1.56	1.53
38......	1.70	20.47	1.70	1.68	1.67	1.62	1.70	1.70	1.70	1.70	1.68
39......	1.79	22.91	1.79	1.79	1.79	1.80	1.86	1.86	1.86	1.86	1.85
40......	2.06	25.65	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.07
41......	2.40	28.72	2.40	2.40	2.40	2.40	2.29	2.29	2.29	2.29	2.33
42......	2.56	31.27	2.56	2.56	2.73	2.73	2.56	2.56	2.56	2.56	2.62
43......	3.06	33.96	3.06	3.06	3.06	3.05	2.87	2.87	2.87	2.87	2.95
44......	3.23	36.39	3.23	3.23	3.23	3.23	3.23	3.23	3.24	3.23	3.31

TABLE 4—*Continued*

Age	Crude Mortality Rate per 1,000	Exposure in Units of 1,000,000	$\theta=2$	$\theta=3$	$\theta=4$	$\theta=5$	$\theta=10$	$\theta=15$	$\theta=20$	$\theta=35$	Published Mortality Rate per 1,000
			Graduated Value								
45......	3.66	38.79	3.66	3.66	3.66	3.66	3.66	3.66	3.66	3.66	3.73
46......	4.34	40.89	4.34	4.34	4.34	4.34	4.15	4.15	4.15	4.15	4.22
47......	4.70	43.23	4.70	4.70	4.70	4.70	4.70	4.70	4.70	4.70	4.74
48......	5.14	45.49	5.14	5.14	5.14	5.14	5.31	5.31	5.32	5.31	5.31
49......	6.14	47.54	6.14	6.14	6.14	6.14	5.99	5.99	5.99	5.99	5.96
50......	6.66	49.39	6.66	6.66	6.66	6.66	6.66	6.66	6.66	6.66	6.61
51......	7.23	50.86	7.23	7.23	7.23	7.23	7.33	7.33	7.31	7.33	7.29
52......	8.05	51.75	8.05	8.05	8.05	8.05	8.05	8.05	8.03	8.05	8.02
53......	8.81	52.44	8.81	8.81	8.81	8.81	8.81	8.81	8.81	8.81	8.82
54......	9.64	52.73	9.64	9.64	9.64	9.64	9.69	9.68	9.69	9.68	9.69
55......	10.73	52.29	10.73	10.73	10.73	10.73	10.69	10.67	10.67	10.65	10.66
56......	11.73	51.50	11.73	11.73	11.73	11.73	11.73	11.73	11.73	11.73	11.73
57......	12.82	50.78	12.82	12.82	12.82	12.82	12.82	12.87	12.87	12.92	12.91
58......	14.20	50.12	14.20	14.20	14.20	14.20	14.20	14.20	14.20	14.23	14.24
59......	15.73	49.57	15.73	15.73	15.73	15.73	15.73	15.73	15.73	15.67	15.71
60......	17.47	47.02	17.47	17.47	17.47	17.47	17.35	17.31	17.31	17.24	17.31
61......	18.93	46.04	18.93	18.93	18.93	18.93	18.93	18.93	18.93	18.93	19.02
62......	20.74	44.76	20.74	20.74	20.74	20.74	20.74	20.74	20.74	20.74	20.90
63......	23.14	43.41	23.14	23.14	23.14	23.14	22.78	22.78	22.78	22.78	22.95
64......	25.03	41.22	25.03	25.03	25.03	25.03	25.03	25.03	25.03	25.03	25.15
65......	27.50	36.75	27.50	27.50	27.50	27.50	27.50	27.50	27.50	27.50	27.52
66......	30.19	34.31	30.19	30.19	30.19	30.19	30.19	30.19	30.19	30.19	30.04
67......	33.47	31.90	33.47	33.47	33.10	33.10	32.67	33.10	33.06	33.06	32.71
68......	34.12	29.33	34.12	34.12	34.12	34.93	34.95	36.22	36.13	36.13	35.56
69......	38.42	27.18	38.42	38.42	38.42	38.42	38.42	39.69	39.37	39.37	38.88
70......	43.58	24.83	43.58	43.58	43.58	43.58	43.07	43.49	42.80	42.80	42.47
71......	46.85	22.36	46.85	46.88	46.88	46.88	46.85	46.85	46.27	46.27	45.99
72......	49.76	20.05	49.76	49.76	49.76	49.76	49.76	49.76	49.76	49.76	49.46
73......	52.23	17.82	52.23	52.23	52.23	52.23	52.43	52.50	53.29	53.29	53.19
74......	56.45	15.62	56.45	56.45	56.73	56.73	56.45	56.45	57.49	57.49	57.58

TABLE 4—*Continued*

Age	Crude Mortality Rate per 1,000	Exposure in Units of 1,000,000	$\theta=2$	$\theta=3$	$\theta=4$	$\theta=5$	$\theta=10$	$\theta=15$	$\theta=20$	$\theta=35$	Published Mortality Rate per 1,000
			Graduated Value								
75......	63.27	13.52	63.27	63.27	63.27	63.27	61.83	61.62	62.37	62.37	62.73
76......	70.00	11.69	70.00	68.79	68.69	68.69	67.75	67.42	67.92	67.92	68.36
77......	72.98	10.00	73.40	73.00	72.98	72.98	74.19	73.86	74.16	74.16	74.35
78......	79.62	8.54	79.62	79.62	79.63	79.63	81.15	80.94	81.06	81.06	81.16
79......	88.65	7.27	88.65	88.65	88.65	88.65	88.65	88.65	88.65	88.65	88.79
80......	99.90	6.09	99.04	99.04	97.85	97.54	96.67	97.00	96.91	96.91	96.72
81......	106.70	5.00	106.70	106.70	106.70	106.31	105.22	105.98	105.85	105.85	104.38
82......	111.64	4.01	111.64	111.64	115.19	114.96	114.30	115.59	115.47	115.47	112.28
83......	116.93	3.23	119.86	119.86	125.20	125.10	124.81	125.84	125.76	125.76	121.68
84......	136.73	2.49	131.36	131.36	136.73	136.73	136.73	136.73	136.73	136.73	133.66
85......	136.64	1.84	146.14	146.14	149.78	149.86	150.08	148.25	148.38	148.38	147.69
86......	170.79	1.41	164.21	164.21	164.35	164.48	164.85	160.41	160.70	160.70	164.30
87......	193.94	1.04	182.64	182.64	180.45	180.60	181.04	173.20	173.70	173.70	180.32
88......	177.85	0.75	201.44	201.44	198.06	198.22	198.65	186.63	187.37	187.37	194.76
89......	220.60	0.55	220.60	220.60	217.20	217.33	217.69	200.69	201.73	201.73	211.14
90......	214.70	0.37	240.13	240.13	237.85	237.93	238.15	215.39	216.76	216.76	227.38
91......	260.03	0.28	260.03	260.03	260.03	260.03	260.03	230.72	232.46	232.46	244.54
92......	285.40	0.19	280.30	280.30	283.73	283.62	283.33	246.69	248.85	248.85	259.43
93......	243.87	0.12	300.93	300.93	308.95	308.71	308.06	263.29	265.91	265.91	274.05
94......	264.48	0.07	321.93	321.93	335.69	335.30	334.21	280.53	283.64	283.64	299.03
95......	367.91	0.05	343.29	343.29	363.95	363.37	261.78	298.40	302.06	302.06	334.98
			Measures of Fit and Smoothness								
$\Sigma\, w_x(u_x''-u_x)^2$			1,783.68	1,799.32	2,273.09	2,284.79	2,419.00	2,629.01	2,513.32	2,515.53	1,681.06
$\Sigma\, w_x\lvert u_x''-u_x\rvert$			115.51	128.63	184.56	211.39	358.45	428.56	466.34	477.15	464.91
$\Sigma\,(\Delta^3 u_x)^2$			199.01	170.84	89.10	66.56	10.97	3.05	0.51	0.46	161.43
$\Sigma\,\lvert\Delta^3 u_x\rvert$			58.73	52.60	37.17	31.68	9.01	3.88	1.56	1.19	35.76

procedures involve linear programming, an operations research tool with which some actuaries may not be familiar but which may very well be useful in areas other than graduation. The method enables the graduator to explore and to define quite precisely the expanse between perfect fit and perfect smoothness, and thereby to select a set of graduated values that strike the proper balance between fit and smoothness. In the case of mortality data over a full range of ages, however, a complete project may not be practicable because of the amount of computer time required to make the calculations for many values of the smoothness parameter θ. Nevertheless, a graduation for a single value of θ may not require any more time than a Whittaker-Henderson graduation and can be performed in any computer facility in which a linear programming subroutine is available. It is hoped that future research will lead to computational shortcuts that will enable the graduator to obtain a complete picture of the range between perfect fit and perfect smoothness even for large data sets.

REFERENCES

1. ARMSTRONG, R. D., and FROME, E. L. "A Comparison of Two Algorithms for Absolute Deviation Curve Fitting," *Journal of the American Statistical Association*, LXXI (June, 1976), 328–30.
2. BARRODALE, I., and ROBERTS, F. D. K. "An Improved Algorithm for Discrete l_1 Linear Approximation," *SIAM Journal of Numerical Analysis*, X (October, 1973), 839–48.
3. DANTZIG, G. B. *Linear Programming and Extension*. Princeton, N.J.: Princeton University Press, 1963.
4. DAVIS, P. J. *Interpolation and Approximation*. New York: Dover Publications, Inc., 1975.
5. GREVILLE, T. N. E. *Part 5 Study Notes—Graduation*. Education and Examination Committee of the Society of Actuaries, 1974.
6. HENDERSON, R. "A New Method of Graduation," *TASA*, XXV (1924), 29–40.
7. HILL, R. W., and HOLLAND, P. W. "Two Robust Alternatives to Least-Squares Regression," *Journal of the American Statistical Association*, LXXII (December, 1977), 828–33.
8. HILLIER, F. S., and LIEBERMAN, G. J. *Operations Research*. 2d ed. San Francisco: Holden-Day, Inc., 1974.
9. MILLER, M. D. *Elements of Graduation*. New York: Actuarial Society of America and American Institute of Actuaries, 1946.
10. RAMSAY, J. O. "A Comparative Study of Several Robust Estimates of Slope, Intercept, and Scale in Linear Regression," *Journal of the American Statistical Association*, LXXII (September, 1977), 608–15.
11. SCHUETTE, D. R. "A Defense of the Karst Algorithm for Finding the Line

of Best Fit under the l_1 Norm," *TSR 1735*, Mathematics Research Center, Madison, Wis., 1977.

12. SIMONNARD, M. *Linear Programming*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1966.
13. WAGNER, H. M. "Linear Programming Techniques for Regression Analysis," *Journal of the American Statistical Association*, L (March, 1959), 206–12.
14. ———. *Principles of Operations Research*. 2d ed. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1975.
15. WHITTAKER, E. T. "On a New Method of Graduation," *Proceedings of the Edinburgh Mathematical Society*, XLI (1923), 63–75.
16. WHITTAKER, E. T., and ROBINSON, G. *The Calculus of Observations*. 4th ed. London and Glasgow: Blackie & Son, Ltd., 1944.

DISCUSSION OF PRECEDING PAPER

JAMES C. HICKMAN AND ROBERT B. MILLER:*

Professor Schuette has provided us with an ingenious formulation of the classical actuarial graduation problem and has shown how the theory of linear programming provides mathematical insights (Theorems 1 and 2) and computational insights (Examples I and II) into this problem. We particularly applaud the illustration of how the use of high-speed electronic computers can yield a "feel" for the data that would not have been feasible even a few years ago. In fact, we feel that Professor Schuette's treatment of the numerical examples is more modern in spirit than his traditional mathematical formulation of the graduation problem.

Specifically, we believe that the $F + \theta S$ model is not an adequate theoretical foundation for graduation as it is actually performed, although the model is of great historical interest. This point is illustrated by the examples in the paper. The modernity of the examples stems from the treatment of the graduations as functions of the smoothing constant θ. Surely this treatment is sensible from a practical point of view, because whatever preconceptions the graduator may have about the smoothness of the true rates can be expressed only in a vague way. This means that he or she will want to inspect a spectrum of graduation until an acceptable one is found. Professor Schuette follows such a plan in Example I, judging a variety of graduations to be "good" or "reasonable" on the basis of their (F, S) values. It is interesting to note that the "objective function" $F + \theta S$ is not used to determine whether a graduation is acceptable. Indeed, some of the graduations deemed acceptable have higher $F + \theta S$ values than those deemed unacceptable.

If θ is fixed, the $F + \theta S$ function can be minimized. However, if one is also permitted to vary θ, the objective function $F + \theta S$ always can be made equal to zero by choosing $u_x = u''_x$ for all x and $\theta = 0$. Thus in fact another implicit "objective function" (function of θ) is being sought in the analysis of the data. Would it not be advisable to make explicit this additional function of F and S by which the graduations are being judged?

From a Bayesian point of view, the $F + \theta S$ model flows from the following assumptions. Let the conditional density of u'', given u, be $f(u'' | u)$, and let the marginal density of u, $p(u | \theta)$, be a function of the parameter θ. Then the posterior density of u, given u'', is proportional to

* Dr. Miller, not a member of the Society, is associate professor of statistics and business, University of Wisconsin—Madison.

$p(u|\theta)f(u''|u)$. If we take

$$f(u''|u) \propto \exp\left(-\sum_{x=1}^{n} w_x |u_x'' - u_x|^q\right), \tag{1}$$

where $q > 0$ and $w_1, \ldots, w_n$ are known, and we take

$$p(u|\theta) \propto \theta \exp\left(-\theta \sum_{x=1}^{n-z} |\Delta^z u_x|^q\right), \tag{2}$$

we obtain the kernel of the posterior density

$$\exp\left[-\left(\sum_{x=1}^{n} w_x |u_x'' - u_x|^q + \theta \sum_{x=1}^{n-z} |\Delta^z u_x|^q\right)\right]. \tag{3}$$

The mode of this posterior density is obtained by minimizing the exponent, so this result contains both Schuette's ($q = 1$) and Whittaker's ($q = 2$) formulations. A number of remarks need to be made at this point.

1. The function $p(u|\theta)$ is a singular density on u, that is, it assigns u to a subspace of n-dimensional space with probability 1. While singularity is not a disastrous quality for a prior distribution, the justification for its use in this case is not immediately apparent. Whittaker [7] may not have thought of $p(u|\theta)$ as a distribution on u at all, but rather in terms of a prior density on S of the form $\theta e^{-\theta S}$, where S could be any of a number of measures of smoothness. Naturally, many other candidates for $p(u|\theta)$ functions come to mind that are at least as reasonable as Whittaker's choice, but his choice has been more or less enshrined in actuarial science.
2. There is no obvious reason why the q in formula (1) should be equal to the q in formula (2). A graduator convinced strongly of the smoothness of the u's might take $q = 4$ in (2) and a lower value of q in (1). There are limiting-distribution arguments supporting the choice of $q = 2$ in (1) (see [3], for example), but robustness considerations might lead one to consider $q = 1$.
3. Robustness considerations are not new. In 1888 Edgeworth [1], following Laplace, considered the minimization of the sum of the absolute values of residuals (the method of situation) as an alternative to least squares. He even noted that such a procedure was less sensitive to outliers than least squares. For his minimization problem he also provided a solution that had a linear programming flavor (see especially the graph used to illustrate the solution).

In our review of the Bayesian approach, we have treated θ as a fixed parameter whose value is chosen a priori by the graduator, and this is the way Whittaker originally treated it. How do we incorporate the idea of a variable θ? One approach is that suggested by Lindley [5] and Lindley and Smith [6] in the context of Bayesian analysis of linear statistical models. Instead of specifying a prior distribution on the vector $\mathbf{u}$, which has a large number of elements, specify a prior distribution on θ with density $h(\theta)$. Then the joint prior density on θ and $\mathbf{u}$ is

$$h(\theta)p(u|\theta) , \tag{4}$$

the posterior density on θ and $\mathbf{u}$ is proportional to

$$h(\theta)p(u|\theta)f(u''|u) , \tag{5}$$

and the posterior density on $\mathbf{u}$ is proportional to

$$\int_0^\infty h(\theta)p(u|\theta)f(u''|u)d\theta . \tag{6}$$

Note that specification of $h(\theta)$ indirectly specifies a prior distribution on $\mathbf{u}$. In this way the dimension of the parameter for which a prior distribution must be specified directly is reduced.

We will illustrate this idea with two examples. If we adopt assumptions (1) and (2) and assume further that θ has a degenerate (one-point) distribution, the traditional result, equation (3), follows. Since Whittaker interpreted θ as the ratio of two variances, it might be plausible to quantify uncertainty about θ in the form of a prior density such as

$$h(\theta) \propto \theta^{\alpha-1}e^{-\beta\theta} , \quad \theta > 0 , \quad \alpha > 0 , \quad \beta > 0 . \tag{7}$$

Then expression (6) is proportional to

$$\exp\left(-\sum_{x=1}^{n} w_x|u_x'' - u_x|^q\right)\Big/\left(\beta + \sum_{x=1}^{n-z} |\Delta^z u_x|^q\right)^{\alpha+1} . \tag{8}$$

Since our knowledge about θ is likely to be vague, $h(\theta)$ typically will be quite diffuse. As $\alpha \to 0$ and $\beta \to 0$, $h(\theta)$ becomes diffuse and expression (8) approaches

$$\exp\left(-\sum_{x=1}^{n} w_x|u_x'' - u_x|^q\right)\Big/\sum_{x=1}^{n-z} |\Delta^z u_x|^q .$$

Maximization of this function leads to minimization of

$$\sum_{x=1}^{n} w_x|u_x'' - u_x|^q + \ln\sum_{x=1}^{n-z} |\Delta^z u_x|^q ,$$

which obviously involves a tradeoff between fit and smoothness, but finding the maximizing u_x presents quite a challenge.

We can sidestep the challenge by remarking that if $h(\theta)$ is very diffuse we can set it equal to 1 for practical purposes and simply maximize

$$p(u|\theta)f(u''|u)$$

with respect to θ and **u**. The resulting **u** would be the vector of graduated values and would have been found by searching over a function of θ. Again using the assumptions (1) and (2), we are led to the maximization of the function

$$\theta \exp\Big[-\Big(\sum_{x=1}^{n} w_x |u_x'' - u_x|^q + \theta \sum_{x=1}^{n-s} |\Delta^s u_x|^q\Big)\Big]. \qquad (11)$$

For fixed θ, the exponent can be minimized and then the value of (11) computed. Repetition of these calculations for various values of θ would lead to an "optimal" set of graduated values.

We close by noting that Kimeldorf and Jones [4] and Hickman and Miller [2] suggested $p(\mathbf{u}|\theta)$ functions that involved multidimensional θ. Such functions make the graduation model more complicated but also much richer, and we believe that the search for realistic $p(\mathbf{u}|\theta)$ functions is the direction that future research in graduation should take.

REFERENCES

1. EDGEWORTH, F. Y. "On a New Method of Reducing Observations Relating to Several Quantities," *Philosophical Magazine*, Ser. 5, XXV (1888), 184.
2. HICKMAN, J. C., and MILLER, R. B. "Notes on Bayesian Graduation," *TSA*, XXIX (1977), 7.
3. HOEM, J. "The Statistical Theory of Demographic Rates," *Scandinavian Journal of Statistics*, III (1976), 169.
4. KIMELDORF, G. S., and JONES, D. A. "Bayesian Graduation," *TSA*, XIX (1967), 66.
5. LINDLEY, D. V. "The Estimation of Many Parameters." In *Foundations of Statistical Inference*, ed. V. P. GODAMBE and D. A. SPROTT (New York: Holt, Rinehart & Winston, 1971), p. 435.
6. LINDLEY, D. V., and SMITH, A. F. M. "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society*, XXXIV (1972), 1.
7. WHITTAKER, E. T., and ROBINSON, G. *The Calculus of Observations.* 4th ed. London and Glasgow: Blackie & Sons, 1944.

STUART A. KLUGMAN:

Professor Schuette is to be applauded for introducing the techniques of linear programming to an actuarial problem. It is hoped that more uses for this powerful tool will be found in the future.

I wish, however, to take exception to the particular application selected by the author. It is clear that the Whittaker-Henderson method does not provide a robust solution to the estimation problem. It appears to me that any robust improvement must attack the fit and smoothness components separately. The fit measure can be generalized as $F = \Sigma\, \rho(s_x)$, where $s_x = (u_x'' - u_x)/\sigma_x$, $\sigma_x = \text{Var}\ (u_x'')$, and $\rho(\ )$ is an arbitrary loss function. If the u_x'' are independent normal random variables, then $\rho(x) = x^2$ provides an "optimal" solution, and with $w_x = 1/\sigma_x^2$ the Whittaker-Henderson Type B formulation results. Schuette has selected a robust alternative, $\rho(x) = |x|$. A compromise covering the range of possibilities between these two approaches was suggested by Huber [1]. He shows that

$$\begin{aligned} \rho(x) &= x^2\,, & |x| &\leq c \\ &= 2c|x| - c^2\,, & |x| &> c \end{aligned}$$

minimizes the maximum asymptotic variance of the estimator, where the maximum is over all random variables s_x of the form $(1 - \epsilon)\Phi + \epsilon\text{H}$, $0 \leq \epsilon \leq 1$; Φ is the standard normal random variable, and H is any symmetric random variable. The parameter c is a decreasing function of ϵ. If $\epsilon = 0$ the normal distribution holds and $c = \infty$, leaving the Whittaker-Henderson form. If $\epsilon = 1$ the class contains all symmetric distributions and $c = 0$. As $c \rightarrow 0$, $\rho(x) \rightarrow |x|$ and Schuette's form results. Huber recommends choosing c between 1 and 2. When applying this method, it turns out that all observations more than c standard deviations from the mean are treated as "outliers" and their contribution to the estimator is reduced. One of the methods of evaluating the estimators is to determine iteratively weights $(v_1, v_2, \ldots)$ such that using $\Sigma\ v_x(s_x)^2$ produces the same solution as that obtained by using $\Sigma\ \rho(s_x)$. This enables each iteration to be performed by the method used for obtaining the least-squares estimator. See [2] for a detailed description of the iterative approach. A major advantage of this method is that the outliers are readily identified by their weights. By selecting c close to zero, it is possible to approximate Schuette's form by this method.

A more serious problem involves the determination of the robust smoothness measure. Smoothing is considered necessary in order to remove sampling fluctuations. The greater the fluctuations, the more smoothing is needed. Suppose a general smoothness measure were $\Sigma\ \nu(\Delta^z u_x)$. It would seem that large values of $\Delta^z u_x$ would indicate those ages at which extra smoothing need be done. Use of a loss function like Huber's ρ would produce the opposite effect, reducing the importance of good smoothness at those ages where it is not good initially. On the smoothness side, the use of $\nu(x) = x^2$ seems very satisfactory, and $\nu(x) = |x|^p$ for $p > 2$ may be even more appropriate.

Table 1 of this discussion presents a graduation of the data used in Example I in Schuette's paper. The function being minimized is $\Sigma\ \rho(s_x) + \theta\Sigma\ (\Delta^3 u_x)^2$ with Huber's ρ function. The variances are estimated as

TABLE 1

A ROBUST GRADUATION OF EXAMPLE I

x	u''_x	Table 2				Robust			
		u_x ($\theta=25.4$)	Δu_x	$\Delta^2 u_x$	$\Delta^3 u_x$	u_x	Δu_x	$\Delta^2 u_x$	$\Delta^3 u_x$
1........	34	22.32				27.33			
			4.36				2.28		
2........	24	26.68		−0.04		29.61		0.57	
			4.32		0.01		2.85		−0.15
3........	31	31.00		−0.03		32.46		0.42	
			4.29		0.01		3.37		−0.01
4........	40	35.29		−0.02		35.83		0.41	
			4.27		−0.02		3.78		−0.08
5........	30	39.56		−0.04		39.61		0.33	
			4.23		0.02		4.11		−0.13
6........	49	43.79		−0.02		43.72		0.20	
			4.21		−0.01		4.31		−0.05
7........	48	48.00		−0.03		48.03		0.15	
			4.18		0.40		4.46		0.02
8........	48	52.18		0.37		52.49		0.17	
			4.55		0.03		4.63		0.00
9........	67	56.73		0.40		57.12		0.17	
			4.95		−0.03		4.80		0.06
10.......	58	61.68		0.37		61.92		0.23	
			5.32		0.02		5.03		0.11
11.......	67	67.00		0.39		66.95		0.34	
			5.71		−0.01		5.37		0.06
12.......	75	72.71		0.38		72.32		0.40	
			6.09		0.00		5.77		0.11
13.......	76	78.80		0.38		78.09		0.51	
			6.47		0.01		6.28		0.04
14.......	76	85.27		0.39		84.37		0.55	
			6.86		−0.01		6.83		0.01
15.......	102	92.13		0.38		91.20		0.56	
			7.24		0.00		7.39		0.03
16.......	100	99.37		0.38		98.59		0.59	
			7.62		0.01		7.98		0.00
17.......	101	106.99		0.39		106.57		0.59	
			8.01		−0.01		8.57		0.01
18.......	115	115.00		0.38		115.14		0.60	
			8.39				9.17		
19.......	134	123.39				124.31			
$\Sigma\ w_x(u''_x-u_x)^2$....		6,557				6,309			
$\Sigma\ w_x\lvert u''_x-u_x\rvert$....		873				883			
$\Sigma\ (\Delta^3 u_x)^2$........		0.17				0.09			
$\Sigma\ \lvert\Delta^3 u_x\rvert$........		0.41				0.95			

$\hat{\sigma}_x^2 = u_x(1 - u_x)/E_x$. Because the weights are different from those used in Schuette's graduation, the appropriate value for θ will be different. A value of $\theta = 2{,}000{,}000$ was used in order to approximate Schuette's results with $13.07 \leq \theta \leq 37.72$. The robust graduation used $c = 1.5$. It will be seen that the two graduations are quite similar. The robust graduation indicated the presence of two possible outliers. A weight of 0.64 was attached to u_6'' and a weight of 0.74 to u_9''. In a general setting this would indicate that the experience at those ages should be examined carefully. The outlying values may be caused by clerical error, the presence of a claim for a large amount, or even a true mortality characteristic that is contrary to the smoothness objective.

A surprising characteristic of Schuette's examples is that values of θ within an interval produce identical graduations. I believe this is connected with the use of the absolute value of the differences. For most mortality data, the differences can be expected to have the same sign over a wide range of ages. In that case their sum telescopes into the difference of lower-order differences at the initial and final ages of this range. Changing the intermediate values will have no effect on S unless the change is large enough to alter a sign in one of the differences. This restricts changes to the ends of such intervals. Such a lack of continuity in the available graduations is not appealing. Examination of the zth differences in the various graduations of Example I shows that the method attempts to make most of the zth differences zero. This leads to good smoothness by the internal definition but often produces one or two large zth differences. For example, with $z = 3$ and $\theta = 25.4$, the Table 2 graduated values consist of two quadratic polynomials, one running from $x = 1$ to $x = 8$ and the other from $x = 8$ to $x = 19$ (see the differences in Table 1 of this discussion).

There is considerable work yet to be done in constructing a robust graduation method. Schuette's paper provides a start in that direction.

REFERENCES

1. HUBER, P. "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, XXXV (1964), 73–101.
2. LENTH, R. "A Computational Procedure for Robust Multiple Regression." University of Iowa Technical Report No. 53, 1976.

HARWOOD ROSSER:

For over a dozen years, whenever I have read a paper on graduation I have found myself wearing two hats. The first one I usually take off, as in this case, in admiration for anyone who can push back the frontiers of knowledge in such a technical area. Wearing the other hat, I ask myself: "Should this topic be added to the required reading and, if so, in what form?"

My answer, for this paper, is: "Not in the present form, at least." It would seem to be a fairly simple idea to replace the squares of deviations, and of certain orders of differences, with absolute values in the graduation process. This project the author considers in the context of Whittaker-Henderson Type B graduation. The difficulties he encounters suggest why this had not been done before.

Absolute values of differences have been suggested before as measures of smoothness. As the author notes, Miller mentions it, in section 1.7 of *Elements of Graduation.* This discussant used absolute values in his 1968 paper "Interpolation by Computer" (*TICA*, Vol. XVIII). The problems arise when a minimization process is attempted.

Intrinsically, as is pointed out, least absolute values are at least sometimes preferable to least squares, as when the error distribution has heavier tails than the normal distribution.

Not being very familiar with linear programming, I shall have few comments on the theory involved. Apparently this technique supersedes the use of orthogonal polynomials, which previously served to reduce the amount of arithmetic in fitting curves of second or higher degree by a least-squares approach.

Examples

As to the author's numerical examples, I shall be more articulate. In his Example I he finds six graduations "acceptable": one in Table 1, two in Table 2, and three in Table 3. I am afraid I am a little more critical, as much of the method as of the results obtained.

With the exception of the right-hand column of Table 3, which is by definition a fitted third-degree polynomial, all of the acceptable graduations show that the lack of smoothness is concentrated at a single point (two points for the Table 1 graduation). This is also true of the "non-acceptable" graduations in Tables 1 and 2, and probably in Table 3 as well (excluding the right-hand columns, which are necessarily polynomials), except that roughness may appear at as many as three points.

This phenomenon takes a form in which the sum of the absolute values of the zth differences is essentially equal to that of a single difference or, at most, of two or three. Thus, in Table 1, the graduated values usually lie on a wide-angled V. For higher orders of differences, the graph of the graduated values would be represented by segments of two different curves, with unequal slopes at the point of junction (cf. Miller, *Elements of Graduation*, p. 20, Fig. 3*a*.) This is the type of situation that osculatory or tangential interpolation was designed to remedy.

A little reflection serves to show how this can happen. If the sum of the absolute values of a certain order of differences is to be taken as the measure of smoothness, this criterion will find the following two situations equivalent, among others: (*a*) one where all differences of that order, except one, equal zero, and (*b*) one where all such differences are nearly equal. Thus, internally, a choice between two or more competing graduations that meet this criterion will be resolved on other grounds, usually that of fit. Obviously, a sum-of-the-squares criterion would find situations *a* and *b* markedly different.

Practical compromises may be possible, however. One would be to use absolute values in connection with measuring fit and to retain the classical sums of squares to measure smoothness. The former would deal with the "outlier" problem and the latter would avoid the problem just described.

Example II (Table 4) does not appear to follow the pattern of Example I. Any comparison with the published values is somewhat distorted, since these were obtained by an unweighted formula that measured smoothness in terms of second differences.

Whittaker-Henderson Type A

Speaking of unweighted formulas, I am disappointed but not surprised that Dr. Schuette makes no mention of a counterpart to the once-popular Type A formula, which can be employed without recourse to a computer. As a step in that direction, I attempted unsuccessfully to find a factorization of the second-order difference equation that would be analogous to Henderson's factorization of the fourth-order equation (cf. Miller, *Elements of Graduation*, sec. 10.7). Hence my lack of surprise.

In the days before electronic computers, Type B calculations were seldom performed, primarily because of the heavy arithmetic involved. Since then, there has been a sharp shift in emphasis from Type A to Type B, as reflected in the required reading. Type A now is often regarded merely as a special case, with all weights equal, of Type B.

I have no quarrel with electronic computers, but I am chary of becoming completely dependent on them. The Education and Examination Committee should beware of this trap. There are still actuaries with small companies or consulting firms who, because of time pressure, expense, or other reasons, may be reluctant to turn to a computer every time a small graduation problem arises. It would be unfortunate if they should acquire the impression that there is no middle ground and that they must resort to graphic graduation.

Even in a large firm, there are questions of waiting in line, priorities, computer down-time, and the like. While time-sharing has improved the situation, these problems have not been eliminated completely. Thus, an intermediate alternative approach sometimes has much to recommend it. This is what a Type A graduation offers.

T. N. E. GREVILLE:

In this paper Donald Schuette has made a most interesting and valuable contribution to the literature on actuarial graduation. He has given cogent reasons why the criterion of least absolute values should be considered as an alternative to least squares and has adapted the Whittaker-Henderson graduation method to the former criterion. He has devised an ingenious computation scheme based on the simplex algorithm of linear programming.

In the introductory section he discusses vector norms and defines the l_p norm of a vector. In this notation, the usual "Euclidean norm" based on squares is called the l_2 norm, while that based on absolute values is called the l_1 norm. The l_p and l_q norms are called "dual" to each other if $1/p + 1/q = 1$. Thus the l_2 norm is dual to itself, while the l_3 and $l_{3/2}$ norms are dual to each other. The dual of the l_1 norm is the so-called l_∞ norm, also called the "uniform norm" or the "Chebyshev norm." The latter may be defined as the maximum of the absolute values of the components of the vector.

The Chebyshev norm plays a very important role in mathematical approximation theory. Thus, a polynomial of degree z that exhibits the best fit in the Chebyshev norm to a set of data points has the property that the maximum absolute deviation is smaller than for any other polynomial of the same degree. Such a polynomial is sometimes called the "minimax" polynomial. In the Whittaker-Henderson adaptation based on this norm, F would be the maximum absolute difference between a graduated value and the corresponding observed value, and S would be the maximum of the absolute values of the zth differences of the graduated values.

It would be most interesting and worthwhile if someone would perform the same task for the Chebyshev norm that Schuette has done for the l_1 norm.

(AUTHOR'S REVIEW OF DISCUSSION)

DONALD R. SCHUETTE:

Professors Hickman and Miller remind us that graduation can be viewed as a statistical estimation problem, and that a Bayesian formulation provides a very general structure that allows for inclusion directly in the model of prior opinion as to smoothness of the underlying values to be estimated and that yields as a special case the $F + \theta S$ objective function for any norm or combination of norms. As they indicate, if the observed values are independently and randomly distributed about the underlying values according to the double exponential distribution, and

$$p(u|\theta) = \theta \exp\left(-\theta \sum_{x=1}^{n-z} |\Delta^z u_x|^q\right),$$

that is, the underlying values u_x have a prior distribution such that $\Sigma\ |\Delta^z u_x|^q$ has the exponential distribution with parameter θ, then the posterior mode is obtained when the graduated values are selected as those that minimize

$$F + \theta S = \sum_{x=1}^{n} w_x |u_x'' - u_x|^q + \theta \sum_{x=1}^{n-z} |\Delta^z u_x|^p .$$

Hickman and Miller remark further that (*a*) setting

$$p(u|\theta) = \theta \exp\left(-\theta \sum_{x=1}^{n-z} |\Delta^z u_x|^p\right)$$

does not specify a complete prior distribution for the underlying values u_x but only specifies a distribution for the univariate function

$$S = \sum_{x=1}^{n-z} |\Delta^z u_x|^p ,$$

(*b*) the norm selected for smoothness does not have to be the same as that selected for fit, and (*c*) suspicions that the square norm ($p = 2$) may not always be the best choice for fit were entertained a century ago.

One point of disagreement that I have with them is over their statement that "a graduator convinced strongly of the smoothness of the u's might take $q = 4$ in (2) and a lower value of q in (1)." It seems to me that the choice of norm for smoothness has more to do with what combinations of values of the u's the graduator considers to be smooth than it does with the degree of his conviction regarding smoothness. A graduator convinced of smoothness, however he defines it, selects a larger value of θ than one who is not so strongly convinced of smoothness relative to fit.

Hickman and Miller reject Whittaker's view that θ in the $F + \theta S$ model is a parameter the value of which the graduator can set in advance. The Bayesian approach, of course, is to set a joint prior distribution on θ and the u's. They offer some suggestions as to how this might be done; some of those suggestions perhaps bear investigation. For example, setting a uniform prior distribution on θ in the $F + \theta S$ model leads, as they indicate, to maximizing their expression (11), which in principle yields a single value of θ and a unique graduation. However, the little experimenting I have performed with this suggestion has not led to any graduations I can accept.

Hickman and Miller give me partial credit for viewing θ as a variable and inspecting graduations over a range of its values. However, they disapprove of starting with $F + \theta S$ as an objective function even though it can be reached from a Bayesian starting point with θ viewed as a parameter. Their preference is for the more complete Bayesian approach to graduation as presented by them and by Kimeldorf and Jones, in which prior opinion concerning more than smoothness may be specified. I will concede that if a graduator has prior opinion beyond notions of smoothness—for example, opinion as to approximate levels of the underlying values themselves—then a full Bayesian approach provides an excellent model for the task. I believe that within the confines of the $F + \theta S$ model my method offers the graduator a practical way of exploring many options between perfect fit and perfect smoothness.

Professor Klugman agrees that the Whittaker-Henderson l_2 norm solution to the graduation problem is not robust. However, he takes issue with my choice of the l_1 norm, especially in the case of smoothness. He points out that using the l_1 norm for smoothness leads to graduations in which many of the zth differences are zero but in which one or two zth differences may be quite large in absolute value. Such a situation violates his notion of smoothness. He suggests using the Huber compromise between the l_1 and l_2 norms for fit and the l_p norm with $p \geq 2$ for smoothness. He describes a procedure for obtaining graduated values when the Huber norm for fit ($c = 1.5$) and the l_2 norm for smoothness have been adopted. The procedure is iterative and at each stage employs the method used for obtaining l_2 norm values. The graduation he obtains is a good one. My questions to him are these: (1) how much experimentation did he have to perform before he decided upon the value $\theta = 2{,}000{,}000$, and (2) how much computer time was required to obtain graduated values for each value of θ?

Professor Klugman expresses some surprise about different values of θ producing identical graduations under the methods of my paper. It is a characteristic of linear programming that one corner point of the feasible region can remain optimal over a range of values of an objective function parameter.

I am not as ready as Professor Klugman to place the label "robust" on the graduated values obtained by his method. I agree that considerable work remains to be done in obtaining robust graduation methods.

Mr. Rosser makes the same point as Klugman regarding the smoothness of the graduated values under my method using the l_1 norm, namely, that the lack of smoothness tends to be concentrated at one or two points. I agree that to anyone brought up under the influence of osculatory interpolation formulas the l_1 norm may not be acceptable for smoothness. Mr. Rosser suggests that the l_1 norm for fit and the l_2 norm for smoothness may produce good results. I agree, although the minimizing problem produced by that combination is not solved readily as far as I know. A linear programming formulation, for example, is no longer possible.

A better solution may lie in following Dr. Greville's suggestion regarding the Chebyshev or uniform norm. The l_1 norm for fit and the Chebyshev norm for smoothness may very well be the combination that satisfies the smoothness objections that have been raised and that copes with the outlier problem. The minimizing problem that arises under that combination of norms can be formulated as a linear programming problem. That problem will receive my attention shortly.

Mr. Rosser expresses concern over becoming completely dependent upon computers to perform graduations. He would like to have available a method that in an emergency could be implemented by hand or desk calculator. I believe that the methods of my paper and perhaps others that are based upon linear programming offer hope to Mr. Rosser. Another project on which I am working is that of developing a heuristic interpretation of each iteration in the linear programming procedure, which may be performable on a desk calculator.

I want to express my appreciation to all the discussants for their interest and advice. The comments made in regard to the choice of norm for smoothness are well taken and perhaps suggest why the method apparently has not been perceived as practical and ready for implementation. Possibly when the smoothness norm difficulty has been overcome, the method of my paper will be viewed differently.