USE OF THE CHI-SQUARE STATISTIC TO SET WHITTAKER-HENDERSON SMOOTHING COEFFICIENTS

Lee Giesecke, Defense Manpower Data Center

ABSTRACT

Whittaker-Henderson graduation has for some time been used by actuaries to smooth crude rate data. The goal of such a graduation is to improve the estimates for the individual points in a series. One of the problems of the Whittaker-Henderson is the amount of subjectivitity that enters into the determination of the degree of smoothness required. This paper suggests a way to use the Chi-square statistic to set the smoothing coefficient when using the Whittaker-Henderson graduation procedure.

1. INTRODUCTION

Actuaries routinely face the problem of estimating a smoothed series of rates from crude rate data. The smoothed rates must estimate the probability of some outcome, such as death, for each point along an axis representing increasing age, length of service, or some other variable. Typically, there is some data for each point, as shown in table 1, but prediction can be improved by making use of the surrounding observations. Techniques for performing such smoothing, or graduation, include moving averages, curve fitting, and graphical methods, as well as the Whittaker-Henderson.

The Whittaker-Henderson is really a venerable technique, it having first been proposed by Whittaker in 1919. Subsequently, Henderson developed a practical procedure for implementing the method.

One aspect of the Whittaker-Henderson that is left to the intuition of the analyst is how smooth the graduated series should be. Obviously, crude rates will be poor predictors unless they are based upon a large number of cases. As increasing smoothness is applied, however, there is a gradual loss of information.

In this paper I want to describe the Whittaker-Henderson technique, to consider whether the Chi-square statistic can be helpful in determining the the degree of smoothness, and to compare results from Whittaker-Henderson graduation with the results of fitting a polynomial to the same data series using regression techniques.

2. WHITTAKER-HENDERSON GRADUATION

Whittaker-Henderson graduation assumes there is a simple trade-off between a measure inversely related to fidelity, F, and a measure inversely related to smoothness, S. The function F + kS is then minimized for a particular value of the smoothing coefficient, which is set by the person performing the graduation.

The fidelity measure is given by

$$F = \sum_{i=1}^{p} w_i \, (u_i'' - u_i)^2$$

Where w is the weight or exposure

u_i'' is the ungraduated rate

u_i is the ungraduated rate

p is the number of points in the series

The measure of smoothness is given by

$$S = \sum_{i=1}^{p} (\Delta^z u_i)^2$$ where Δ^z is a difference operator or order z

For example, $\Delta^1 u_1 = u_2 - u_1$,

$\Delta^2 u_1 = \Delta^1(\Delta^1 u_1) = \Delta^1(u_2 - u_1) = u_3 - 2u_2 - u_1$

In matrix notation S is defined as S = u'K'Ku, where u is the vector of graduated rates and K is a differencing matrix. For example, where n=7 and z=2, K would be given as

$$K = \begin{matrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{matrix}$$

Note that smoothness is measured on the graduated rates, rather than the crude rates. An example of second differences is also shown in table 1.

It is shown by (Greville 1974, p. 53) that F + kS is minimized where

$$(W + kK'K)\, u = Wu''$$

where W is the square matrix with weights w_i, on the main diagonal and u and u" are respectively the vectors of graduated and ungraduated rates.

Greville (1974) also shows that the matrix W + kK'K is positive definite, so that u may be solved for in the above equation using Choleski factorization. The principle problem left to the analyst is to choose the smoothing coefficient.

3. CHI-SQUARE

Usually statisticians use the upper tail of a Chi-square statistic to see whether a test statistic is larger than might be expected (at some probability level) under their null hypothesis. Here we are interested in both tails, since we also want to be able to test whether the graduated rates fit the crude rates <u>more</u> closely than than might be expected, for a particular value of the smoothing coefficientz. Further, we want to see what happens if we iteratively reset the smoothing coefficient, k, until the probability of receiving a higher value of the Chi-square (or a lower value) is sufficiently close to .5.

The Chi-square is obtained from a sum of independent normal random variables with mean 0 and variance 1. We require that each of the p observations in the series u" is based on enough cases, w_i, so that it may be approximated using a normal distribution.

The most straight forward Chi-square occurs where the population probabilities are known. Where v_i is the expected value of u_i'' the expression

$$\sum \frac{w_i \, (u_i'' - v_i)^2}{v_i \, (1 - v_i)}$$

results in a Chi-square with p degrees of freedom (see, for example, Freund 1962). Using matrix notation, the same result is given by $(u''-v)' \Sigma^{-1} (u''-v)$. Here Σ is the square matrix with diagonal elements $v_i\,(1-v_i)/\,w_i$.

Since v is unknown, our problem is to construct a Chi-square using u"-u. From the Whittaker-Henderson solution Au = Wu", it may be noted that $u''-u = (I-A^{-1}W)u''$. Defining $B = (I-A^{-1}W)$, then $u''-u = Bu''$.

Since B is not of full rank the n-vector Bu" has some observations which are linear combinations of the others. To get a valid Chi-square statistic it is necessary to pre-multiply B by P, a reducing matrix of order p-z by p (where z is the order of differencing used).

This is accomplished with $P=P_1P_2$. Here P_2 is a p by p matrix of full rank, which sweeps out elements of B below a trial diagonal. The trial diagonal begins on the main diagonal but shifts one position to the right whenever a linear dependence is found. This will occur, in the sweep out process, when the current trial diagonal element and elements below it are zero. P_1 is an p-z by p matrix with 1's on the main diagonal and zeroes elsewhere.

If u" is approximately distributed as $N_p(v,\Sigma)$, then the linear transformation PB(u"-v) is distributed as $N_{p-z}(0, PB\Sigma B'P')$. (See Timm, 1975, p. 121).

Timm (1975, p.132) gives the very general result, that in this situation $[PB(u''-v)]'\ D\ [PB(u''-v)]$ is distributed as a Chi-square with degrees of freedom equal to the rank of D, if and only if the product $D(PB\Sigma B'P')D = D$ or $D(PB\Sigma B'P')$ is idempotent.

To achieve this result we set $D=(PB\Sigma B'P')^{-1}$, and the resulting Chi-square has degrees of freedom equal to the rank of B.

The rank of B may be obtained as a by-product of the computer program that finds P. This turns out to be less than full rank by the order of difference used in the definition of smoothness. This result is intuitively appealing when one considers that the Whittaker-Henderson preserves a number of moments equal to the order of differences used. For third differences, for example, it may be seen that:

$$[111\ldots\ldots1]Wu = [111\ldots\ldots1]Wu''$$

$$[i_1 i_2 \ldots i_p]Wu = [i_1 i_2 \ldots i_p]Wu''$$

$$[i^2,\ i_2^2 . i_p^2]Wu = [i_1^2\ i_2^2\ ..i_p^2]Wu''$$

This occurs because we set

$$(W + kK'K)u = Wu'' \text{ and}$$

$$[111\ldots\ldots1]K' = [i_1 i_2 \ldots i_p]K' = [i_1^2\ i_2^2\ \ldots i_p^2]K' = 0.$$

Consequently, if the vector u is given and any p-z of the u_i'', we may solve for the remaining z.

4. WEIGHTED LEAST SQUARES REGRESSION

To provide a further basis for comparison weighted least squares regression was also used. The basic model was

$$Y_x = b_0 + b_1x + b_2x^2 + b_3x^3 + b_5x^5$$

where x is the age or length of service and Y_x is the rate at x

However, terms which were not significant were not added. The stepwise procedure in SAS (Goodnight 1979) was used, with the MINR algorithm. This algorithm finds the best 1 variable solution, then adds the independent variable which makes the minimum R^2 improvement. Next, MINR keeps substituting a variable not being used as an independent variable for one which is, selecting the substitution which makes the minimum improvement in R^2. When the best 2 variable solution is found a third variable is added and so forth.

The MINR considers a good many models without performing all 63 possible. The analyst then usually finds a model with high R^2, which also has the desirable property that terms are significant at a fairly high level.

The stepwise procedure in SAS does not incidentally provide a weighted least squares option as such. It can be accomplished, however, by specifying no intercept, creating a dummy variable of 1's, and multiplying all terms, including the dummy variable and the dependent variable, by the square root of the weighting factor. In this case an initial set of regressions was performed using the exposures w_i as the weighting factor and a second set, using $w_i/q_i(1-q_i)$, where q_i is the predicted rate at i using the equation from the initial run. It may be seen that this process satisfies the asusmptions of weighted least squares (Draper and Smith, 1966)

5. CROSS VALIDATION

The results from the Whittaker-Henderson graduation and the regression were used to predict 10 different rates. The predictors are based on grouped-data for the fiscal years 1977-79 and were used to predict fiscal 1980. Comparisons giving the proportion of explained 1980 variance, R^2, are given in table 2.

The left-most column of table 2 shows the rate being compared. Beside this is given the R-squared when the ungraduated 1977-79 rates were used to predict the 1980 rates. It should be noted that for comparisons 5, 6, 7 and 8 the ungraduated rates did as well as the Whittaker-Henderson graduations and in comparisons 6 and 8 did better than the regression estimates.

The two Whittaker-Henderson graduations were nearly equal in their performance. In the fourth comparison the intuitively set Whittaker-Henderson had a higher cross-validation R^2. In the ninth the Chi-squared set Whittaker-Henderson did better. Otherwise the results are nearly equal.

Except for comparisons 6, 8 and 9 the regression based estimates had as high an R^2 as the Whittaker-Henderson. In the fourth comparison the regression estimates had a higher R^2.

It should be noted that with the definition of R^2 used here, that it is possible for the R^2 to be negative. This did occur in a few cases, where the 1980 values were not widely dispersed and the 1980 mean was closer to the 1980 rates than were the 1977-79 predictions. The definition of R^2 is given at the bottom of table 1.

6. DISCUSSION

6.1 ANTICIPATED USE OF WHITTAKER-HENDERSON WITH CHI-SQUARE.

In theory a Whittaker-Henderson solution using an analyst-set smoothing coefficient can yield better predictions than a Whittaker-Henderson where the smoothing coefficient is iteratively set to yield a 50th percentile Chi-square. If, by chance, an ungraduated series is already smooth, the Chi-square iterations would enhance peaks and valleys more than they should be. If by chance, the ungraduated series is ragged, the 50th percentile graduation may not be smooth enough.

It is remarkable then that 50% percentile graduation does as well as the graduation with intuitively set coefficients. At the very least we can say that 50th percentile solution provides a good initial solution. If a more refined solution is needed, one might proceed as follows:

1. Arrive at the solution for the 50th percentile Chi-square.
2. Bracket the solution in 1, with solutions using higher and lower Chi-square levels.
3. Examine the results of 2 noting changes in the measures of fidelity and smoothness as well as the Chi-square level.
4. Proceed in the direction of the preferred solution until a satisfactory solution is found.

6.2 WHITTAKER-HENDERSON vs. REGRESSION

It will come as no surprise to statisticians that regression can do a credible job of smoothing, but it may come as a surprise to actuaries. Although some actuaries are ardent advocates of regression (Tenenbein and Vanderhoof), graduation is still more commonly used. The question that needs to be discussed is where one should use regression and where one should use graduation.

First, it must be noted that when the correct model is known, regression will give minimum variance unbiased estimates. These are nice properties, and they are not merely academic. In some cases the Whittaker-Henderson will do some very strange things in the tails of distributions where data are sparse enough to have little impact on the fidelity measure but where the smoothness measure is operating at full force. Interestingly, if we look at the results in table 1, it will be seen that the only comparisons where regression did not do as well or better was comparisons 6 through 9. In all these comparisons the ungraduated rates were already good predictors.

On the other hand, it seems foolish to throw away much of the information in a rate series based upon a large number of observations. The Whittaker-Henderson will in most cases help prediction and at least will do no harm.

This suggests that a reasonable policy would be to use regression where data are sparse and to use the Whittaker-Henderson where the number of cases is large.

6.3 PROBLEMS AND REFINEMENTS

The most serious problem with the Chi-square use here is the amount of computer time required. One solution to this is to simply use the ordinary Chi-square, substituting

$$u_i \text{ for } v_i \text{ in } \sum \frac{w_i\,(u_i''-v_i)^2}{v_i\,(1-v_i)},$$

but still using the reduced degrees of freedom. Doing this greatly reduced the computer time and gave results almost identical to those as shown in table 2.

More conclusive results may be obtained using simulations rather than the 10 comparisons summarized in table 2. We, of course, cannot simulate the judgment of an analyst, but some simulation results will be useful, nevertheless.

Two further refinements that will also be interesting are the use of logarithms for the rates and the use of alternative definitions of smoothness. Certain of the rates are nearly linear in the logarithms, and there is no reason why the logarithmic smoothing cannot be done. Likewise, there are many possible definitions of smoothness other than the sum of the squares of differences used here.

REFERENCES

Draper, N.R. and Smith, H. (1966) Applied Regression Analysis, John Wiley and Sons

Freund, John E. (1962) Mathematical Statistics, Prentice-Hall, Inc.

Goodnight, James H. et. al. (1979) SAS Users Guide, SAS Institute, Inc.

Greville, T. N. E. (1974), Part 5 Study Notes, Graduation, Education and Examination Committee of the Society of Actuaries

Miller, Morton D. (1949), Elements of Graduation, monograph published jointly by the Actuarial society of America and the American Institute of Actuaries.

Tenenbein, Aaron and Vanderhoof, Irwin T. (1980) "New Mathematical Laws of Select and Ultimate Mortality", Transactions of the Society of Actuaries, 32

Timm, Neil H. (1975), Multivariate Aanlysis with Applications in Education and Psychology, Wadsworth Publishing Co.

Table 1

Ungraduated and Graduated
Enlisted Death Rates

Age	Number of Cases	Crude Rates	Smoothed Rates	First Differences	Second Differences
17	15,821	.0021	.0018		
				-.0001	
18	206,990	.0017	.0017		.0000
				-.0001	
19	452,647	.0016	.0016		.0000
				-.0001	
20	550,073	.0015	.0015		.0000
				-.0001	
21	511,732	.0014	.0014		.0000
				-.0001	
22	415,869	.0012	.0013		.0000
				-.0001	
23	335,864	.0011	.0012		.0000
				-.0001	
24	283,507	.0009	.0011		.0000
				-.0001	
25	241,594	.0010	.0010		.0000
				-.0001	
26	203,932	.0009	.0009		.0001
				.0000	
27	173,906	.0010	.0009		.0000
				.0000	
28	154,407	.0008	.0009		.0000
				.0000	
29	142,204	.0007	.0008		.0000
				.0000	
30	133,289	.0009	.0008		

Table 2

Comparison of Graduation Techniques

Cross Validation R^2 by Graduation Technique

Rate	Ungraduated '77-'79 Rates	Whittaker-Henderson With Smoothing Coefficient Set Intuitively	by Chi-square	Weighted Least Squares Regression
1. Enlistee Death Rate (ages 17-60)	.18	.22[2/]	.22[2/]	.22 intercept, x, x^2
2. Officer Death Rate (ages 19-60)	-.05	.25[2/]	.23[2/]	.25 x, x^2, x^3, x^4
3. Regular Enlistee Permanent Disability Rate (Length of service 0-16)	.53	.58[3/]	.59[3/]	.59 intercept, x, x^2, x^3
4. Reserve Officer Permanent Disability Rate (Lengths of service 0-16)	.08	.22[2/]	.16[2/]	.29 intercept, x^2
5. Reserve Officer Retirement Rate (lengths of service 19-34)	.91	.90[3/]	.91[3/]	.91 intercept, x, x^2
6. Regular Officer Retirement Rate (Lengths of service 19-27)	.33	.32[3/]	.33[3/]	.19 intercept, x, x^3
7. Regular Officer Other Loss Rate (Lengths of service 2-11)	.95	.95[3/]	.95[3/]	.94 x, x^2, x^4, x^5
8. Reserve Officer Other Loss Rate (Lengths of service 4-18)	.80	.80[3/]	.81[3/]	.77 intercept, x
9. Regular Enlistee Temporary Disability Rate (20-37 years of service)	-.56	-.60[2/]	-.51[2/]	-1.00 intercept, x, x^2
10. Regular Officer Temporary Disability Rate (19 - 39 years of service)	-.05	.01[2/]	.00[2/]	.01 intercept, x^2

1/ $R^2 = 1 - \dfrac{\sum w_i (u''_i - u_i)^2}{\sum w_i (u''_i - \bar{u}'')^2}$ Where w_i is the '80 exposure, u''_i the '80 ungraduated rate, u_i is the graduated '77-79 rate or regression predication, and $\bar{u}''$ is the mean '80 rate.

2/ Using second differences
3/ Using third differences
4/ The terms used in the regression equation are indicated in the right hand column.