Volume Title: The Smoothing of Time Series

Volume Author/Editor: Frederick R. Macaulay

Volume Publisher: NBER

Volume ISBN: 0-87014-018-3

Volume URL: http://www.nber.org/books/maca31-1

Publication Date: 1931

Chapter Title: Third-Degree Parabolic Graduation Formulas Having Smooth Weight Diagrams, Summation Formulas

Chapter Author: Frederick R. Macaulay

Chapter URL: http://www.nber.org/chapters/c9363

Chapter pages in book: (p. 51 - 62)

CHAPTER III

Third-Degree Parabolic Graduation Formulas Having Smooth Weight Diagrams. Summation Formulas.

The actuaries have attacked the problem of obtaining sets of weights which lend themselves to easy computation and which, when fitted to second (or third) degree parabolas, will fall exactly on those parabolas but which, when fitted to irregular, non-mathematical data, will give smoother results than can be obtained from the weights which give the mid ordinate of a second-degree parabola fitted by the method of least squares.[1] A number of specialized formulas, each of which was designed to be applied to a specified number of terms of the data, have been developed by different investigators. Spencer's 21-term formula may be used as an illustration. The weights in that formula are:

$$-\frac{1}{350},\quad -\frac{3}{350},\quad -\frac{5}{350},\quad -\frac{5}{350},\quad -\frac{2}{350},\quad +\frac{6}{350},\quad +\frac{18}{350},$$

$$+\frac{33}{350},\quad +\frac{47}{350},\quad +\frac{57}{350},\quad +\frac{60}{350},\quad +\frac{57}{350},\quad +\frac{47}{350},\quad +\frac{33}{350},$$

$$+\frac{18}{350},\quad +\frac{6}{350},\quad -\frac{2}{350},\quad -\frac{5}{350},\quad -\frac{5}{350},\quad -\frac{3}{350},\quad -\frac{1}{350}.$$

[1] See Robert Henderson and H. N. Sheppard, *Graduation of Mortality and Other Tables*, the files of the Journal of the Institute of Actuaries, and the Transactions of The Actuarial Society of America,

Such a set of weights gives excellent results which are quite easy to compute. If the data were a monthly time series, the computer would first take a 7-months weighted moving total of the data with the following simple set of weights: −1, 0, +1, +2, +1, 0, −1. He then would take a 5-months moving total of a 5-months moving total of a 7-months moving total of the results. The final figures would then each be divided by 350.[1]

Such a formula as Spencer's 21-term formula tends to give a smooth graduation because the

[1] Note that $350 = 2 \times 7 \times 5 \times 5$. If superlative smoothness is not insisted upon, it is easy to discover a set of averages which will give a weight diagram such that when applied to a second (or third) degree parabola the results will fall on that parabola. The procedure may be illustrated from Spencer's 21-term formula above. Having picked the original set of weights, −1, 0, +1, +2, +1, 0, −1, on the basis of general judgment, the investigator calculates where a curve based upon such a set of weights would come if fitted to the second-degree parabola $y = x^2$. Multiplying the above set of weights by +9, +4, +1, 0, +1, +4, +9, he gets a total of −16 which divided by the total of the weights gives a figure of −8. In other words, this fitted curve will fall 8 units *outside* the parabola $y = x^2$. However, a simple 7-months moving average of the same parabolic data will fall 4 units *inside* the parabola and a simple 5-months moving average 2 units *inside* the parabola. If, after taking the original average, he continues by taking a 7-months moving average of the results and two 5-months moving averages, he raises the results of the first fitting by +4, +2, +2, or +8, bringing the fitted curve back on the parabola. A simple moving average of n consecutive terms comes $\frac{(n^2 - 1)}{12}$ C *inside* the parabola $y = A + Bx + Cx^2$.

A 21-term formula, which is often a desirable substitute for Spencer's, may be applied as follows: Take a 9-months weighted moving total of the data with the following simple set of weights:

weight diagram is itself a smooth curve. Recognition of the fact that the smoothness of the resulting graduation depends directly on the smoothness of the weight diagram led Robert Henderson to propose a solution of the problem of obtaining a general expression for the smoothest possible weight diagrams which, when fitted to second (or third) degree parabolas, would fall exactly on those parabolas. Assuming that lack of smoothness could be measured by calculating the sum of the squares of the third differences of the weights in the weight diagram, he developed a general formula which would make the sum of the squares of these third differences a minimum for any number

—1, 0, +1, +1, +1, +1, +1, 0, —1. Take a 3-months moving total of a 5-months moving total of a 7-months moving total of the results. Divide each of the final figures by 315.

The graduation resulting from the use of the above formula will tend to be a shade smoother than that resulting from the use of Spencer's 21-term formula. It will tend to fit cyclical data more adequately (it falls ⅔ *outside* the parabola $y = x^2$). It is a shade simpler to compute.

Another 21-term formula which may be substituted for Spencer's is the following: Take a 9-months moving total of the data with the same weights as before, namely: —1, 0, +1, +1, +1, +1, +1, 0, — 1. Take a 3-months moving total of a 4-months moving total of an 8-months moving total of the results. Divide each of the final figures by 288. This formula is almost exactly parabolic (⅙ outside $y = x^2$). It gives a weight diagram distinctly smoother than Spencer's 21-term formula. It fits short period sine curves somewhat less adequately than Spencer's formula. This is because it contains an 8-months moving average while Spencer's formula contains no average longer than 7. Like the preceding formula, it is a shade less laborious to apply than Spencer's 21-term formula.

of terms desired.[1] If the number of terms desired in a formula be represented by 2m − 3, Robert Henderson shows that the general expression for the xth term is

$$\frac{315\{(m-1)^2-x^2\}\{m^2-x^2\}\{(m+1)^2-x^2\}\{(3m^2-16)-11x^2\}}{8m\,(m^2-1)\,(4m^2-1)\,(4m^2-9)\,(4m^2-25)}{}^{2}$$

To derive a set of 15 weights from the above general formula, 9 is substituted for the letter m and the values of the above expression are calculated for each value of x from −7 to +7. Making

[1] The successive orders of "differences" equal either the deviation or a multiple of the deviation of a datum point from a parabola put through as many consecutive data points as possible, excluding the particular datum point whose deviation is measured.

The successive differences are the *actual* deviations of a datum point from parabolas put through as many consecutive data points as possible which immediately *precede* the datum point under discussion. Thus, the nth difference is the deviation of a datum point from an (n−1)th degree parabola. For example, if the ordinates of five successive data points be 7, 17, 43, 91, 190, and if through the first four of these ordinates we put a third-degree parabola, its equation will be $y = 7 - 3x + 2x^2 + x^3$. Now the fifth ordinate given by this equation is 167. The deviation of the fifth data ordinate from the fifth ordinate of the curve is therefore +23. But this is the fourth difference of the five data ordinates.

The successive differences are *multiples* of the deviation of a datum point from a parabola fitted to other than immediately preceding data points. For example, let there be 8 consecutive data points. Through the first five data points and the last two data points put a sixth degree parabola. Then the deviation of the sixth datum point from this parabola will equal $1/21$ of the seventh difference of the eight data points. Notice that the seventh difference is $-y_1 + 7y_2 - 21y_3 + 35y_4 - 35y_5 + 21y_6 - 7y_7 + y_8$.

[2] Robert Henderson and H. N. Sheppard, *Graduation of Mortality and Other Tables*, p. 35; and Transactions of The Actuarial Society of America, Vol. XVII, p. 43.

the necessary substitutions in the formula, we obtain the following set of weights for a 15-term weight diagram:

$$-\frac{2652}{193154}, \quad -\frac{3732}{193154}, \quad -\frac{2730}{193154}, \quad +\frac{3641}{193154}, \quad +\frac{16016}{193154},$$
$$+\frac{28182}{193154}, \quad +\frac{37422}{193154}, \quad +\frac{40860}{193154}, \quad +\frac{37422}{193154}, \quad +\frac{28182}{193154},$$
$$+\frac{16016}{193154}, \quad +\frac{3641}{193154}, \quad -\frac{2730}{193154}, \quad -\frac{3732}{193154}, \quad -\frac{2652}{193154}.$$

The total, of course, equals unity.

The reader might be interested in comparing this "Ideal" formula with Spencer's 15-term formula.[1] Spencer's formula for 15 terms gives the following set of weights:

$$-\frac{3}{320}, \quad -\frac{6}{320}, \quad -\frac{5}{320}, \quad +\frac{3}{320}, \quad +\frac{21}{320}, \quad +\frac{46}{320}, \quad +\frac{67}{320},$$
$$+\frac{74}{320}, \quad +\frac{67}{320}, \quad +\frac{46}{320}, \quad +\frac{21}{320}, \quad +\frac{3}{320}, \quad -\frac{5}{320}, \quad -\frac{6}{320},$$
$$-\frac{3}{320}.$$

The sum of the squares of the third differences of these weights is only 12 per cent greater than the corresponding sum in Henderson's Ideal formula.[2]

[1] Spencer's 15-term formula is applied by taking a 5-months weighted moving total of the data with the following simple set of weights: -3, $+3$, $+4$, $+3$, -3. A 4-months moving total of a 4-months moving total of a 5-months moving total of the results are then taken. The final figures are each divided by 320. The actual computation of a Spencer 15-term graduation is given in Appendix II of this book.

[2] A 15-term strictly third-degree parabolic formula which is even better than Spencer's 15-term formula may be applied as fol-

This is an extremely low price to pay for the computation advantages of such a summation formula as Spencer's.[1]

Examining second-degree parabolic summation formulas and Henderson's Ideal formulas in the light of the four reasons already given for not using a graduation based on the mid ordinates of second-degree parabolas fitted by the method of least squares, we find that the first objection still holds. Such parabolic summation formulas do not eliminate seasonal fluctuation except by accident. However, they may be so constructed as to do so. All that is necessary to eliminate a 12-months sea-

lows. Take a 5-months weighted moving total of the data with the following simple set of weights: − 10, + 11, + 10, + 11, − 10. Take a 3-months moving total of a 4-months moving total of a 6-months moving total of the results. Divide the final figures by 864. The sum of the squares of the third differences is slightly less than in Spencer's formula. The shape of the weight diagram is distinctly closer to a Henderson's Ideal weight diagram. The formula is slightly less laborious to apply than Spencer's 15-term formula.

[1] The sum of the squares of the third differences of a set of weights which give the mid ordinate of a second-degree parabola fitted to 15 observations by the method of least squares is more than 26 times the corresponding sum in Henderson's Ideal formula. However, though this is bad, it is not quite as alarming as it sounds. The sum of the squares of the third differences of a graduation which has been applied to data by a least squares parabolic formula will not necessarily be 26 times the sum of the squares of the third differences of a graduation resulting from the use of the Henderson formula; in particular cases it may be very little greater. Unless many decimals are carried it will seldom be more than 3 or 4 times as great. Smoothness of the weight diagram, though important, is only one element in smoothness of the graduation.

sonal is to have a 12-months moving average as part of the calculation scheme. Henderson's Ideal formula cannot contain a 12-months moving average in its computation scheme as it is not computable by moving averages. It can therefore entirely eliminate seasonal fluctuation only by accident.[1]

The second objection to the use of weight diagrams based on the mid ordinates of second-degree parabolas fitted by the method of least squares is that such a method of fitting does not result in the smoothest possible curves. The third objection is based on the disadvantage of second (or third) degree parabolic fitting to periodic functions. Corresponding to these two objections is a sort of compound objection against the use of second-degree parabolic summation formulas or Henderson's Ideal formula. As parabolas are not periodic functions, if more than a small number of terms are used in a second-degree parabolic summation formula, the resulting graduation of any periodic type of data will tend to be unmistakably too low at maximum points and too high at minimum

[1] If a correct number of terms be used in a Henderson Ideal formula, it will practically eliminate seasonal fluctuations. For example, if the seasonal fluctuation be a 12-months sine curve, a Henderson 33-term Ideal formula will eliminate 101½ per cent of the seasonal. This little 1½ per cent over-elimination is, of course, generally negligible. A 25-term Ideal formula will eliminate only 76 per cent of such a sine seasonal, while a 37-term Ideal formula will eliminate 105 per cent of the seasonal.

points. However, if such a small number of terms be used in the formula as to overcome this difficulty, we are faced with the fact that in such series as monthly Call Money Rates the number of terms in the formula will not then be great enough to attain a high degree of smoothness.

Furthermore, it is not possible to have a smooth and well-shaped parabolic weight diagram which contains a 12-months moving average unless there are at least 27 terms in the diagram.[1] The second column of the table in Appendix III gives the weights for the smoothest possible 25-term second-degree parabolic graduation formula which will eliminate 12-months seasonal fluctuations. An examination of that table and Figures 7, 15, and 16 of Chart I will show the reader the inevitable hollow in the middle of the weight diagram, which comes in all 12-months weight diagrams which have an insufficient number of terms. The sum of the squares of the third differences of such an "ideal" 25-term, 12-months seasonal-eliminating set of weights is almost four times as great as in

[1] Even if such a summation formula as the Kenchington 27-term formula did not undesirably dampen the underlying cycles in such a series as Call Money Rates, it would not be usable if we insisted upon *rigid* elimination of 12-months seasonal fluctuations. If the seasonal fluctuation be a 12-months sine curve, Kenchington's formula will eliminate over 90 per cent of the seasonal. Spencer's 21-term formula will eliminate less than 45 per cent of such a seasonal. For Kenchington's 27-term formula, see note 3, page 29; for Spencer's 21-term formula, see pages 51, 52, 53.

the 25-term Henderson Ideal set of weights (which have no seasonal-eliminating condition).

For the investigator who wishes to use a 12-months seasonal-eliminating second-degree parabolic weight diagram (instead of such a diagram as Kenchington's 27-term formula, which does not entirely eliminate seasonal fluctuations), an excellent 29-term formula is the following: Take a 4-months moving total of an 8-months moving total of the data. Subtract a 17-months moving total of the data. Take a 2-months moving total of a 12-months moving total of the results. Divide by 360.[1] Such a formula, if fitted to the parabola $y = x^2$, gives results falling 7/6 *outside* the parabola. The case would hardly ever occur in practice where this extremely small amount outside the parabola would not be an advantage rather than a disadvantage in fitting. The weight diagram is smooth and well shaped. The 29 weights are:

$$-\frac{1}{360},\ -\frac{3}{360},\ -\frac{5}{360},\ -\frac{6}{360},\ -\frac{5}{360},\ -\frac{2}{360},\ +\frac{3}{360},$$

$$+\frac{9}{360},\ +\frac{15}{360},\ +\frac{21}{360},\ +\frac{27}{360},\ +\frac{32}{360},\ +\frac{36}{360},\ +\frac{39}{360},$$

$$+\frac{40}{360},\ +\frac{39}{360},\ +\frac{36}{360},\ +\frac{32}{360},\ +\frac{27}{360},\ +\frac{21}{360},\ +\frac{15}{360},$$

[1] This formula may also be applied as follows: Take a 15-months moving total of the data with the following simple weights: −1, 0, 0, 0, +1, +1, +1, +1, +1, +1, +1, 0, 0, 0, −1. Take a 2 of a 3 of a 12-months moving total of the results. Divide by 360.

$$+\frac{9}{360}, \quad +\frac{3}{360}, \quad -\frac{2}{360}, \quad -\frac{5}{360}, \quad -\frac{6}{360}, \quad -\frac{5}{360}, \quad -\frac{3}{360},$$
$$-\frac{1}{360}.$$[1]

However, such a 29-term parabolic formula will not reach up to the maximum values of the underlying curve of such a series as monthly Call Money Rates, or down to the minimum values. Better results can often be obtained if we do not restrict ourselves to a curve which must exactly fit a second-degree parabola. For example, take such a formula as $\frac{1}{432}$ of a 3-months moving total of a 12-months moving total of the result of subtracting a 16-months moving total of the data from a 4-months moving total of a 7-months moving total of the data.[2] This formula contains 29 terms. That the weight diagram is not particularly close to the ideal form is not important. It falls 3½ outside the second-degree parabola $y = x^2$. It is fitted to 97 months of Call Money Rates in column 14 of the table in Appendix VIII.[3]

[1] This 29-term approximately third-degree parabolic formula is used in the study of interest rates and security prices to obtain a graduation for *long time trend.* For that purpose the formula was applied to the Januaries, Mays and Septembers of the 43-term graduation. Intermediate values were read off a large scale chart on which a smooth curve had been drawn through these values by means of French curves. This procedure gave results quite accurate enough for our purposes and, of course, involved comparatively little work.

[2] See pages 26 and 27 and Figure 14, Chart I.

[3] See column 14 of the table in Appendix IV for the weights calculated to 5 decimals.

The fourth objection to the use of formulas giving the mid ordinates of second-degree parabolas fitted by the method of least squares is that the weight diagrams do not lend themselves to easy computation. Of course, this objection does not hold against the parabolic summation formulas but it holds in its fullness against Henderson's Ideal formulas.

Before concluding this chapter, it is desirable to draw the reader's attention to a relation between graduation by the mid ordinates of third-degree parabolas fitted by the method of least squares and graduation by third-degree parabolic formulas having smooth weight diagrams.

In graduation by the mid ordinates of third-degree parabolas fitted by the method of least squares, the assumption is made that the deviations of the data from the fitted parabola are given equal weights. On this assumption, the sum of the squares of the deviations is made a minimum. Graduation by third-degree parabolic formulas having smooth weight diagrams may be described as graduation by the mid ordinates of third-degree parabolas fitted by the method of least squares—when the various deviations of the data from the fitted parabola are given certain weights. The sum of the squares of the deviations of the data from the fitted parabola—when these deviations are given these specific weights—is made a minimum.

For example, if a third-degree parabola be fitted to 17 consecutive observations by the method of least squares, and if the seventeen deviations of the data from the curve be given the following positive weights— $\frac{1}{39}$, $\frac{1}{12}$, $\frac{2}{11}$, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{5}{8}$, $\frac{6}{7}$, 1, 1, 1, $\frac{6}{7}$, $\frac{5}{8}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{2}{11}$, $\frac{1}{12}$, $\frac{1}{39}$, the mid ordinate of the parabola fitted by the method of least squares will be the same value as obtained by applying Higham's[1] smoothing formula to the data.[2]

Graduation by Henderson's Ideal formula (see page 54) gives the same results as would be obtained by graduation by the mid ordinate of a third-degree parabola fitted to $2m - 3$ observations by the method of least squares, if the weights, assigned to the successive deviations, *themselves* gave the smoothest[3] possible weight diagram. The formula for the relative weights assigned to the deviations in the case of Henderson's Ideal formula, is:

$$W_x = \{(m-1)^2 - x^2\}\{m^2 - x^2\}\{(m+1)^2 - x^2\}$$

The total of the $2m - 3$ weights, of course, equals:

$$\frac{2}{35}(m^2 - 1)(4m^2 - 1)(4m^2 - 9)$$

[1] See note 1, page 133.

[2] Compare Robert Henderson's *Note on Graduation by Adjusted Average*. Transactions of the Actuarial Society of America, Vol. XVII, p. 45.

[3] Sum of the squares of the third differences a minimum.