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: The Extension of Summation Formulas to Cover the Range of the Data

Chapter Author: Frederick R. Macaulay

Chapter URL: http://www.nber.org/chapters/c9368

Chapter pages in book: (p. 113 - 117)

CHAPTER VIII

The Extension of Summation Formulas to Cover the Range of the Data.

None of the graduations described in this book —except the Whittaker-Henderson graduations and the Rhodes curves—extends the entire length of the data. Extension of the graduation is always necessary. For example, if a 43-term graduation is to be made to cover the entire range of the data, 21 months at each end must be fitted by some other means. The case for extension of the graduation by means of hypothetical extrapolated data, outside the range of the actual data given, has already been briefly stated.[1]

If the investigator does not wish to use hypothetical data, simple mathematical methods of extrapolating the graduation without the use of further data are, of course, available. For example, a 43-term graduation may be made to cover the last 21 months by the following procedure: To the last 43 data points, fit a third-degree parabola by the method of least squares, in such a manner that it will have its middle, or 22nd, point coincident with the last point of the 43-term graduated curve,

[1] See pages 25, 26, 95, and 96, and Henderson, *Graduation of Mortality and Other Tables*, pages 37, 38 and 39.

and that its slope at this 22nd point will be the same as the slope of the 43-term curve at this last point of the 43-term curve. The slope of the 43-term curve at this last point is assumed to be the slope, at this point, of a second-degree parabola passing through the last 3 points of the 43-term curve.[1] It is better to fit a third-degree than a fourth-degree parabola. The chance of getting an absurd termination to the graduation is much greater with a fourth-degree parabola than with a third-degree.

[1] If the ordinates of the last 3 points of the 43-term curve be designated y_{-2}, y_{-1}, y_0, the slope at y_0 of the *second*-degree parabola $y = a + bx + cx^2$ drawn through these 3 points will be

$$\frac{y_{-2} - 4y_{-1} + 3y_0}{2}$$

If now a *third*-degree parabola ($y = A + Bx + Cx^2 + Dx^3$) be fitted to the last 43 data points by the method of least squares in such a manner that it goes through the last point of the 43-term curve, and has at this point the slope of the above fitted *second*-degree parabola at this point, the value of its 4 constants will be as follows:

$$A = y_0$$

$$B = \frac{y_{-2} - 4y_{-1} + 3y_0}{2}$$

$$C = \frac{\Sigma x^2 Y - y_0 \, \Sigma x^2}{\Sigma x^4}$$

$$D = \frac{\Sigma x^3 Y - B \, \Sigma x^4}{\Sigma x^6}$$

where y_{-2}, y_{-1}, y_0 are points on the 43-term curve and Y is an observational value (not a point on a smooth curve). The x's are, of course, points along the time axis. x_0 is the x of the 22nd point.

If the data contain a pronounced seasonal fluctuation, that fluctuation should be estimated by some good method (see Appendix I) and eliminated from the last 43 points of the data before the third-degree parabola is fitted to those data points. The extrapolation will then not be affected by mere seasonal elements in the data.

Such purely mathematical extrapolation is easier to defend than any method which involves judgment. The construction of hypothetical data, of course, always involves some judgment. However, the investigator must remember that it will not make a vast difference to the 43-term curve where the 21 extrapolated data points are placed, so long as they lie within a reasonable range.

The method is much less dangerous than any merely freehand extrapolation of the graduated curve itself. In a sense, the graduation is applied to data which at most are not quite half extrapolated. The procedure may be thought of as semi-mathematical. However, even a mere freehand extension of the graduation itself is not quite so bad as it sounds. The probable error of any fitted curve or graduation which covers the entire range of the data is bound to be large at the ends. A forecast of the future is inevitably implicit in the last few terms of the graduation. A friend of the writer once expressed himself on this matter. Said he, "I should certainly be willing to eat my hat if I could not forecast the future of Call Money Rates on the New York Stock Exchange more accurately than a third-degree parabola."

The forecast implicit in the mathematical extrapolation is bad as often as good. Indeed, the nature of a third-degree parabola is such that it seems as though the forecast were bad more often than good. If the last month for which we have observations be a panic month, such as November 1907, extension of the graduation by means of a third-degree parabola gives a curve rising more and more sharply, in other words, forecasting Call Money Rates at the end of 1908 as a large number of times the height they were in the high month of 1907. A third-degree parabola fitted to Railroad Stock Prices for the 43 months ending August 1909 would give a picture of an advancing market with no sign of hesitation. Of course, in this case, a hypothetical extension of the data might have been little better. The purely mathematical method may be used when the reader feels incapable of making any guess whatsoever. It can hardly then give worse results than introducing hypothetical data.

The argument in favor of purely mathematical methods of graduation which is derived from the fact that they are in general less laborious to apply than good judgment methods does not hold in the case of extending the graduation. The amount of computing necessary to cover the last points on the data by means of a fitted third-degree parabola is much greater than the amount necessary to extend the graduation by applying a summation formula to hypothetical data.