Volume Title: Outlay and Income in the United States, 1921-1938

Volume Author/Editor: Harold Barger

Volume Publisher: NBER

Volume ISBN: 0-870-14159-0

Volume URL: http://www.nber.org/books/barg42-2

Publication Date: 1942

Chapter Title: Appendix G: Note On Statistical Methods

Chapter Author: Harold Barger

Chapter URL: http://www.nber.org/chapters/c5490

Chapter pages in book: (p. 369 - 373)

# Appendix G

## NOTE ON STATISTICAL METHODS

### §1. *Graduation*

Reference has been made on a number of occasions to the use of graduation formulae, particularly for reducing annual to quarterly data. These formulae have to be employed for two main purposes. (a) Where interpolating media are available, but are subject to bias, the raising factors must be graduated in order to secure comparability over year ends (see Chapter I, §3). (b) Where no interpolating media are available, graduation has been used for the quarterly distribution of the annual totals themselves.

For the first purpose a straight line was usually adequate. For the second purpose, however, a moving cubic graduation was frequently adopted. The results yielded by the two methods are similar, but the moving cubic, although a little more trouble to compute, has the advantage that it adds up more nearly to the annual data at turning points. For this reason it was usually preferred. Both methods will be briefly described.

A. *Straight Line Interpolation by Moving Average*. If $A$, $B$, $C$ . . . are annual figures, and if the quarterly figures to be obtained are identified by suffix, then:

$$\begin{aligned} A_3 &= (7A + B)/32 \\ A_4 &= (5A + 3B)/32 \\ B_1 &= (3A + 5B)/32 \\ B_2 &= (A + 7B)/32 \end{aligned}$$

etc.

This represents a properly centered moving average, computed from an even quarterly distribution of the original data, and of course yields a series of straight lines. It has the disadvantage that two initial and two final quarters have to be derived by linear extrapolation, or in some other way. In the case of the first two quarters of 1921, however, data for 1920 were frequently available. The main disadvantage of this sort of interpolation is common to all moving averages—its amplitude at turning points is too small. For this reason the total for the four quarters of any year may depart appreciably, in the neighborhood of such points, from the annual total; i.e.,

$$(A + 6B + C)\ /8$$

need not equal $B$. As already pointed out, rigid agreement between $(B_1 + B_2 + B_3 + B_4)$ and $B$ during the process of interpolation was not insisted upon. Nevertheless it is possible to overcome these last two disadvantages by graduating the annual series

$$\begin{gathered} 8B/(A + 6B + C), \\ 8C/(B + 6C + D), \\ \text{etc.,} \end{gathered}$$

and multiplying the result by the original graduation of the data. The first term in the final graduation which can be obtained by this method is

$$B_3 = \frac{7B + C}{32}\left\{\frac{7B}{A + 6B + C} + \frac{C}{B + 6C + D}\right\}$$

and the other terms are similar. This refinement, which can be applied more than once, increases the amplitude of the graduation at turning points in the original series, and makes the final graduation sum up more nearly, year by year, to the annual figures. It is really equivalent to treating the original straight line graduation as an interpolating series, and then graduating the raising factors. But the application of such successive approximations is rather laborious, and, as explained in Chapter I, §3, has rarely been considered worth while in the present study.

B. *Moving Cubic.* With this method the annual figures are supposed to be centered in the middle of the year, and a cubic is fitted to the points in sets of four. Quarterly figures are read off the cubic for the four quarters lying in each case between the centers of the two middle years to which the curve is fitted. This arrangement conforms better at turning points, and adds up more nearly to the original figures, than does a straight line graduation. Moreover, thanks to W. L. Stevens,[1] an extremely simple method of computation is available. A weighted average of the four values is taken, the weights being as follows:

[1] "Integration and Interpolation," *Annals of Eugenics*, VIII, Part IV (1938), pp. 387–401. I am grateful to R. A. Fisher for drawing my attention to Stevens' work.

| | | | | |
|---|---|---|---|---|
| 3rd quarter | −35, | 945, | 135, | −21 |
| 4th " | −65, | 715, | 429, | −55 |
| 1st " | −55, | 429, | 715, | −65 |
| 2nd " | −21, | 135, | 945, | −35 |

The sum of the weights for each quarter is 1,024. In deriving quarterly from annual figures it is of course necessary to divide by 4,096. This method is scarcely more trouble to compute than a straight line, and was fairly extensively used in the present study. Its only disadvantage is that six items are missing at each end, and have to be supplied in some other way. It can be used with curves of other degree, and with panels of quite different arrangement from that illustrated here. It seems a pity indeed that the method is not more widely known among economists, for in a great many cases it might be substituted advantageously for the moving average and for other less satisfactory graduations.

## §2. *Seasonal Variation*

The various quarterly estimates presented in this volume are shown, so far as possible, both before and after seasonal adjustment. This plan allows the reader to observe the extent, and (if he wishes) to test the adequacy, of the adjustments made.[2] The final totals for outlay and income (to be found in Tables 11 and 18 respectively) have of course been adjusted throughout to exclude seasonal movements. The measurement of seasonal variation has given rise to a considerable body of literature,[3] and only a few points, of special interest to the present study, will be discussed here.

The basic method of seasonal adjustment employed is as follows. Ratios of trend values (obtained either by moving average or moving cubic) to the unadjusted data were computed and set out for each quarter in the form of an array. A series of means was then derived for each of the four arrays by excluding successive pairs of extreme observations, and the tendency of this series noted. The measure of central tendency obtained as a result of this procedure is essentially a

[2] In several important fields, of which the estimates for residual income and for inventory changes and profits are outstanding, seasonally unadjusted data could not be presented, on account of the character of the underlying material or of the methods employed in manipulating it. In this connection see especially Appendices B and C respectively.

[3] Cf. Simon Kuznets, *Seasonal Variations in Industry and Trade* (National Bureau of Economic Research, 1933); A. Wald, *Berechnung und Ausschaltung von Saisonschwankungen* (Vienna, 1936); Horst Mendershausen, "Methods of Computing and Eliminating Changing Seasonal Fluctuations," *Econometrica* (July 1937) 5, pp. 234–62.

mean of the central half or two thirds of the observations. The four means so obtained are next adjusted if necessary so that their sum shall equal 4. The original data are then multiplied by these ratios in order to remove the seasonal variation.

In the case of series such as those for residual income and for inventory changes, which are sometimes positive and sometimes negative, the above procedure has of course to be modified through the use of differences instead of ratios, but is otherwise the same. Where a series fluctuates very widely, as a percentage of its mean, but nonetheless remains positive for all observations, either through logical necessity or because of the character of the period studied, there is sometimes doubt as to the character of the adjustment—whether by differences or by ratios—which is most appropriate. (An example is the series for residual income in Public Utilities which might, but never does, become negative.) Where a series represents essentially the difference between two sets of data, for example gross income and expenses or successive end-of-quarter inventories, there may be real doubt as to whether the operation of seasonal influences upon the residual should be regarded as mainly arithmetic or mainly geometric in character. Probably the best way to correct such a series for seasonal variation is to apply ratio adjustments to the two master series from which it comes as a residual. In some cases it has been possible to do this, but in others no such procedure is feasible, and a choice between the arithmetic and the geometric adjustment has perforce to be made; the residual series is then corrected as it stands.

The seasonal movement in a number of series appears to vary with time; frequently, for example, it disappears altogether during the depression. An adjustment which varies through time can be fixed up, or the series may be divided into two parts. The first alternative was regarded with suspicion, for it is extremely apt to lead to rationalizations, on the part of the computer, which have no basis in fact. Nevertheless it had to be employed to correct a number of the net income series for individual corporations discussed in Appendix B. The second alternative, when it can be applied, appears to be preferable, for the element of judgment involved is much more restricted. It was used, for instance, in adjusting the series for the net cash deficit of the Federal government for seasonal variation.

The seasonal adjustments actually made might perhaps be improved through the expenditure of further labor. It is possible also that the author's prejudice against adjustments which vary through time may not be shared by other students. The residual income and the outlay series for Construction are probably the most difficult to adjust seasonally, at any rate among the more important constituents of the totals, and it is not claimed that the adjustment to either is entirely satisfactory.