ERROR AND UNRELIABILITY IN SEASONALS

By

EDGAR Z. PALMER

An article in the *Annals of Mathematical Statistics* for February, 1930, entitled "A Mathematical Theory of Seasonals," by the Statistical Department of the Detroit Edison Company, has three objects. It presents a mathematical version of the time series analysis, suggests the "interpolation" method of computing seasonals, and constructs a theoretical time series as a test of the new method. The mathematical analysis and the theoretical series are based upon the assumption that the trend, cycle, and seasonal are proportional to each other, while the "errors" or residuals are additive in nature. The reasoning is not necessarily valid for series where the cycle or the seasonal is additive rather than proportional to the trend.

The interpolation method as proposed consists in (1) finding the total of the items for each of the twelve months, and (2) dividing each total by a function which theoretically contains the trend and the cycle insofar as they influence the particular month. In practice this twelve-month function turns out to be a smooth trend curve, and the method of its calculation inspires little confidence that it can reflect much cyclical influence. The function for each month is simply a weighted sum of the annual totals, the weights varying for different months. The early years are weighted more heavily in finding the values of the function which apply to the first half of the year, while the later years are given a greater weight in the second half of the year. The function is influenced almost solely by trend, or rather, by the difference between the first year and the last year of the data, since these two years are the only ones whose weights vary considerably from month to month. It is certain that no cyclical movement, however violent, can have the proper effect upon this function unless it affects the two extreme years.

Since the interpolation method involves dividing the monthly totals (for which may be substituted the monthly means) by a function which is mainly composed of trend, we are justified in considering it a variation of the well-known monthly-means method.[1] In the theoretical series of the Detroit Edison article, the means of each month, corrected for trend by the Davies method, yield a seasonal index almost identical with that obtained by the interpolation method (see Table I). It should be noted that we used the very easily computed semi-means line to correct the monthly means for trend. The semi-means trend in this series is not as steep as the theoretical trend used in the con-

TABLE I

SEASONALS OF THE THEORETICAL SERIES

As Obtaind by Five Methods

Month	Theoretical index*	True index	1 Monthly means	2 Interpolation*	3 Link relative*	4 Ratio to trend-cycle Moving mean	5 Ratio to trend-cycle With free-hand correction
January	.990	.978	.931	.938	.975	.978	.989
February	.930	.908	.880	.885	.890	.918	.905
March	1.050	1.020	.984	.988	.988	1.003	1.037
April	1.020	1.028	1.018	1.021	1.007	1.051	1.035
May	1.040	1.063	1.062	1.065	1.030	1.069	1.064
June	.980	.969	.986	.986	.962	.982	.957
July	.980	.978	.994	.993	.972	.986	.976
August	1.000	1.007	1.019	1.017	1.013	1.024	1.016
September	.980	1.009	1.031	1.028	1.033	1.004	.995
October	1.040	1.056	1.084	1.079	1.099	1.049	1.055
November	.990	.974	.990	.985	1.004	.955	.958
December	1.000	1.013	1.016	1.009	1.027	.982	1.014

* From the Detroit Edison article.

1. Davies, *Economic Statistics* (1922), p. 117.

struction of the series. For the purposes of this quick and easy seasonals method, however, the semi-means trend is accurate enough.

Any one series used in the comparison of methods should, of course, be viewed merely as an illustration, or at most as a sample, of their results. A thousand such series, constructed upon a thousand variations in assumptions, is necessary for determinative comparisons. Long and short series, large and small seasonals, cycles, trends, and irregulars, regular and irregular seasonals, curved and straight trends, additive and proportional combination of the factors: each of these attributes introduces some elements of error into the computation of sasonals, and the errors are not necessarily constant as between different methods.

An instance of the danger of using any single theoretical series occurs in the Detroit Edison article. If we test the residual factor, we find that it also contains some seasonal variation. The true seasonal index of the series, then, is the theoretical seasonal modified by whatever seasonal is to be found in the residuals. There is, as might be expected, some seasonal inequality in the cyclical factor as well, but since it is the task of the method used to eliminate this cyclical influence, we do not consider it a part of the true seasonal. No method, however, can be expected to distinguish between a seasonal arbitrarily designated as the theoretical, and one which is added as part of the residual factor. In Table I, the first column gives the theoretical, and the second column the true seasonal, of the series.

The authors compare their results with those by the link-relative method, and find that the interpolation method gives an index slightly closer to the theoretical seasonal. For further test, we have computed the seasonal by the somewhat more logical ratio-to-trend-cycle method. We used the twelve-month moving mean, centered on the sixth month, to represent the combined trend and cyclical factors. Then we found the ratios of each monthly item in the series to the corresponding moving mean figure, and, after arraying the ratios for each of the twelve months, we found the modified median of each array. This is approximately the method suggested by the Federal Reserve Board,[1] and gives a seasonal with very much less error than the previously mentioned

1. Joy and Thomas, The Use of Moving Averages in the Measurement of Seasonal Variations, *Journal of the American Statistical Association*, vol. 23, p. 241.

methods. (See Table II)

The ratios to the moving mean may be expected to be free from both trend and cyclical influence. However, the moving mean has its faults, especially in its tendency to cut corners when the true cycle makes a sharp change of direction, and in its failure to extend to the ends of the series. For this reason we made another computation, using a corrected moving mean. The data and the moving mean were graphed together, and a free-hand curve was drawn (without reference to the theoretical trend-cycle curve given in the article), correcting the moving mean in three places and extending it to the limits of the series. The seasonal index computed from the ratios to the new trend-cycle curve had an error about half that of the interpolation method. (See Table II)

TABLE II

ERROR OF SEASONALS

Method	(1) Mean deviation from theoretical seasonal	(2) Mean deviation from true seasonal	(3) Mean deviation from 1.0000	(4) Ratio of (2) to (3)
Theoretical seasonal	——	.0158	.0258	.612
True seasonal	.0158	——	.0324	——
Monthly means	.0291	.0197	.0388	.508
Interpolation	.0269*	.0170	.0378	.450
Link relative	.0277*	.0198	.0355	.558
Ratio to trend-cycle:				
Moving mean	.0208	.0132	.0344	.384
With free-hand correction	.0164	.0078	.0328	.239

* From the Detroit Edison article.

When W. I. King[1] first proposed the use of a free-hand curve in

1. King, An Improved Method for Measuring the Seasonal Factor, *Journal of the American Statistical Association*, vol. 19, p. 301.

this connection, objection was raised that this introduced the personal equation into what should be mechanically determinable. In many series, however, any experienced statistician would draw a curve which would fit the data better than the moving mean. There is not as much discretion involved in drawing the curve as there is in choosing between two mechanical methods. The possible error due to the personal factor is very much less than the error made certain by the use of any more mechanical method.

An important test of the reliability of the methods of finding seasonals, which may be applied to actual series where the true seasonal is unknown, consists in an examination of the monthly arrays. The monthly-means and the interpolation methods depend for their reliability upon the distribution of the arrays of the original data from the means of each month. Similarly, the link-relative method depends upon the scatter of the link relatives about their medians, and the ratio-to-trend-cycle methods upon the arrays of ratios. We measured the dispersion for each month for any method by the mean deviation of its array about its central tendency. This should be divided by the central tendency itself to obtain a relative dispersion measure for that month. Then the mean of all twelve dispersion measures was taken,

TABLE III

UNRELIABILITY OF SEASONALS

Method	Relative mean deviation of monthly arrays
Monthly means	.1838
Interpolation	.1838
Link relative	.0734
Ratio to trend-cycle:	
Moving mean	.0621
With free-hand correction	.0492

to give an indication of the unreliability of the method as a whole. The great unreliability of any method based on the monthly means is apparent from Table III, as well as the superiority of the ratio-to-trend-cycle method with free-hand correction.

Some question may arise concerning the propriety of submitting the link relatives to this test of reliability, because the manipulations to which the medians are subjected before they emerge as a seasonal index may decrease the error inherent in the spread of the monthly arrays. We have not been able to derive the algebraic relationship between the mean deviation of the link relatives and the corresponding unreliability of the final seasonal indexes based upon them. In erratic series, the process of computing link relatives tends to heighten the spreading effects of rapid changes in direction; conversely, cumulative multiplication of the median link relatives possibly decreases the error. If so, the link-relative method is not as unreliable as Table III would seem to show it.

The penalty which the computer pays for accuracy and reliability is, of course, a longer time of computation. The time required for the application of each method to the given series is shown in Table IV. The time allowed is for each operation to be performed twice, and the results checked against each other. In addition to the five methods used throughout this article, the time is given for a short cut to the best method, involving much more of the personal equation. The short

TABLE IV

COMPUTING TIME OF SEASONALS

Method	Time in minutes
Monthly means	60
Interpolation	110
Link relative	160
Ratio to trend-cycle:	
Moving mean	285
With free-hand correction	495
All free-hand curve	371

cut consists in not computing the moving mean at all, but drawing the trend-cycle curve altogether free-hand.

This timing, of course. assumes that the seasonal index is the whole objest of the computation. If we were finding the seasonal only as a part of a general statistical analysis of the series, the seasonal should not be charged with the full time for the steps which are useful for other parts of the analysis. The trend-cycle curve, for instance, has other uses than in computing seasonals. In considering the question of speed, it should also be recognized that the various methods do not have the same relative time for series of different length, nor when more elaborate calculating equipment is available than we used.

High speed of computation is not as necessary in seasonals as it is, shall we say, in index numbers. The calculation of a seasonal is a task that does not have to be repeated often for any one series. It is more in the nature of a capital expenditure than a current routine. For student theses, and for investigations where the computation of seasonals is merely incidental or can be roughly done, the monthly-means method is adequate. But for a positive study of actual seasonal influences, and for the elimination of the seasonal factor from indexes published currently by research bureaux, the best method should be used regardless of the longer time needed.

Edgar Z. Palmer