MODIFICATIONS OF THE LINK RELATIVE AND INTERPOLATION METHODS OF DETERMINING SEASONAL VARIATION

By

RICHARD A. ROBB

In a recent paper[1] the statistical department of the Detroit Edison Company have introduced a new method of calculating seasonal variation in a time series. Briefly, the time series ${}_{\bullet}u_x$ is represented by the function ${}_{\bullet}u_x = f(x)\ c(x)\ s(x) + \varepsilon_x$ where $f(x)$ represents secular trend, $c(x)$ cycle, $s(x)$ seasonal, and ε_x residual errors, and by the Method of Least Squares the seasonal variation for any one month will be given by

$$(A) \qquad s_i = \frac{\Sigma\ {}_{\bullet}u_x\ f(x) \cdot c(x)}{\Sigma\left[f(x)\ c(x)\right]^2} \qquad i = 1, 2, 3, \ldots 12.$$

where $s(i)$ represents the seasonal variation in the i th month and the summations in the right hand member of the equation are taken over the years covered by the time series.

If the Method of Moments be used

$$(B) \qquad s_i = \frac{\Sigma\ {}_{\bullet}u_x}{\Sigma\left[f(x)\ c(x)\right]}$$

The trouble lies in the determination of the denominator $\Sigma\left[f(x)\ c(x)\right]^2$ or $\Sigma\left[f(x) \cdot c(x)\right]$. The Detroit Edison have overcome this difficulty by smoothing the observed time series with a sixth degree parabola, keeping the total population for each year unchanged over a period of seven years. In this way seasonal

1. A Mathematical Theory of Seasonals, *Annals of Math. Stat.*, I, p. 57.

variation is obtained from (A) or (B), (B) being much easier to handle than (A).

There appears to be an objection in fitting a curve over a period of seven years and thus for successive seven year intervals obtaining smoothed values for a time series of any given length. The ordinates of the smoothed curve are not equally weighted as, for example, in fitting curves over a ten-year period, the first smoothed ordinate for a year is given by *one* curve, the second by *two* curves, the third by *three* curves, the fourth, fifth, sixth and seventh by *four* curves, the eighth by *three,* the ninth by *two* and the tenth by *one.* To overcome this I decided that my smoothed curve should have the same zero, first and second moments as the observed curve over a period of twelve months. This simply means that a parabola of 2nd degree was fitted to the successive twelve month intervals, and as above a smoothed curve will be obtained for any length of time.

If the observed values u_x are plotted against the corresponding values of x and a parabola of second degree fitted to the points

$$u_{-n},\ u_{-n+1},\ \cdots,\ u_0,\ \cdots,\ u_n$$

determining the constants by the Method of Least Squares, the ordinate of the curve at $x = 0$ is taken as the graduated value of u_0. If $n = 6$ this would involve thirteen observed ordinates, whereas I desire twelve. This difficulty, however, is easily removed by finding a first approximation to my graduated value by using thirteen ordinates; having found the corresponding seasonal variation by (A) or (B), the thirteenth ordinate is divided by this seasonal factor. The parabola which is to represent the smoothed curve given by trend $\times$ cycle is then found from the twelve ordinates subject to seasonal, trend and cycle influences, and a thirteenth from which seasonal has been eliminated.

The graduated ordinate at $x = 0$ corresponding to u_0 is (first approximation)

$$\text{(C)}\qquad u_0' = \frac{1}{143}\Big[25u_0 + 24(u_1 + u_{-1}) + 21(u_2 + u_{-2}) + 16(u_3 + u_{-3}) + 9(u_4 + u_{-4}) - 11(u_{-6} + u_6)\Big] - [c]$$

For example, if we take thirteen ordinates, commencing at January, 1904, and finishing at January, 1905, the first approximation for July is

$$\text{July} = \frac{1}{143}\left[\begin{array}{l} -11\ (\text{Jan.})\ 1904 + 9\ (\text{Mar.}) + 16\ (\text{Apr.}) + 21\ (\text{May}) \\ +24\ (\text{June}) + 25\ \ (\text{July}) + 24\ (\text{Aug.}) + 21\ (\text{Sept.}) \\ +16\ (\text{Oct.}) + 9\ (\text{Nov.})\ -11\ (\text{Jan.})\ 1905 \end{array}\right]$$

where (1) I have designated the production for any one month by the corresponding name of the month, and (2) the formula is rearranged in a form suitable for the calculating machine.

If formula (B) is used it is readily seen that to obtain the seasonal variation for any month we must

(1) Sum together all the Januaries, then all the Februaries, etc. It should be noted that, as the first six months and the last six months of a time teries are not weighted equally with the others, no graduated points were found for these periods. In consequence, as will be seen in practice, two sets of summations of the different months are required, the first including every year except the last, and the second excluding the first year. Then apply formula (C). This gives

$$\Sigma\left[f(x)\quad c(x)\right]$$

(2). Divide $\Sigma_{\bullet} u_x$ by $\Sigma\left[f(x)\quad c(x)\right]$

Having obtained a first approximation to the seasonal factors, $\Sigma\left[f(x)\quad c(x)\right]$ is recomputed as explained above. In practice this is quickly executed, as will be seen in an example completely worked out below.

To illustrate this method I have taken the theoretical time series given by the Detroit Edison. Summing the productions for the various months, we have Table I.

To find the seasonal for *July*, for example, we have to find the value of

$$\Sigma[f(x)\cdot c(x)] = \frac{1}{143}\left[\begin{array}{llll} -11(20434) & + 9(21621) & +16(22615) & +21(23035) \\ +24(21129) & +25(21508) & +24(22118) & +21(22212) \\ +16(23186) & + 9(21215) & -11(21215) & \end{array}\right]$$

Using formula (B) the seasonal for July is $s_7 = \frac{21508}{22280}$ 0.965

TABLE I

Month	1904-1914	1905-1915	1st approx.	2nd approx.
January	20,434	21,215	.971	.973
February	19,425	20,143	.918	.919
March	21,621	22,389	1.015	1.014
April	22,615	23,196	1.045	1.039
May	23,035	24,231	1.061	1.062
June	21,129	22,567	.974	.974
July	21,508	22,820	.965	.967
August	22,118	23,077	.987	.993
September	22,212	23,707	1.011	1.010
October	23,186	24,964	1.071	1.067
November	21,215	22,712	.987	.982
December	21,836	23,182	1.002	1.005
			1200.07	1200.05

In this way the seasonals in Column 4 of Table I were obtained.

The second approximation is obtained with little extra trouble; for July, on account of the thirteenth ordinate, in this case the January of the following year, which has a seasonal of .971, we have to replace the last term in 143 $\Sigma\left[f(x) \quad c(x)\right]$ given above by $\frac{-11(21215)}{.971}$, i. e., the recomputed $\Sigma\left[f(x) \quad c(x)\right]$ is now $\frac{3186016 - 11(21215)(0.0299)}{143}$, the reciprocal of 0.971 being 1.0299.

The seasonals obtained with these corrections are given in Column 5 of Table I.

Comparing the seasonals with actual values, we have the following table.

	Jan.	Feb.	March	April	May	June
Actual Seasonal	.990	.930	1.050	1.020	1.040	.980
Computed Seasonal	.973	.919	1.014	1.039	1.062	.974
Error	–.017	–.011	–.036	+.019	+.022	–.006
	July	**Aug.**	**Sept.**	**Oct.**	**Nov.**	**Dec.**
Actual Seasonal	.980	1.000	.980	1.040	.990	1.000
Computed Seasonal	.967	.993	1.010	1.067	.982	1.005
Error	–.013	–.007	+.030	+.027	–.008	+.005

The mean and standard deviations are compared with the Interpolation Method of the Detroit Edison.

	Mean Deviation of Errors	Standard Deviation of Errors
	.0168	.0194
Interpolation Method	.0269	.0337

It will be noticed that the new method of smoothing yields a standard deviation which is roughly a little greater than half that obtained by the Interpolation Method.

To test whether any actual difference in Seasonals would be obtained by using formula (A), the ordinates of the smoothed curve were found by formula (C) and are given in Table II. The seasonals were as follows:

Jan.	Feb.	March	April	May	June	
.967	.988	1.009	1.072	.986	1.005	
July	**Aug.**	**Sept.**	**Oct.**	**Nov.**	**Dec.**	**Total**
.971	.906	1.007	1.042	1.062	.980	11.995

figures practically identical with those previously obtained.

In Figure I, I have plotted against the various months (1) the actual Seasonal Indices, (2) those given by the Detroit Edison Inter-

TABLE II

	1904	1905	1906	1907	1908	1909
January		1724	1916	2105	1306	1656
February		1765	1968	2092	1180	1765
March		1824	1955	2041	1121	1889
April		1855	1956	2049	1141	2007
May		1885	1956	2093	1187	2129
June		1854	1954	2148	1218	2216
July	1405	1880	2019	2127	1223	2310
August	1490	1882	2063	2064	1253	2370
September	1522	1859	2078	1917	1307	2458
October	1569	1867	2099	1772	1376	2462
November	1634	1890	2122	1618	1465	2502
December	1709	1939	2139	1454	1566	2538

	1910	1911	1912	1913	1914	1915
January	2565	1919	2157	2521	2233	1699
February	2574	1872	2222	2562	2123	1803
March	2571	1861	2293	2590	2022	1920
April	2544	1852	2370	2592	1900	2057
May	2529	1869	2441	2609	1863	2248
June	2461	1900	2514	2645	1821	2449
July	2382	1935	2540	2640	1774	
August	2275	1972	2554	2624	1735	
September	2167	1993	2536	2511	1645	
October	2036	2008	2488	2405	1634	
November	1950	2048	2477	2314	1586	
December	1931	2113	2464	2271	1613	

polation Method, and (3) those given by the method of this paper.

As it will be interesting to note how the smoothed values of the ordinates of the time series agree with the actual, I have given below the Mean Deviation of errors from actual for the various months.

Month	Jan.	Feb.	March	April	May	June
M. D.	58	41	50	43	32	34
Month	July	Aug.	Sept.	Oct.	Nov.	Dec.
M. D.	23	24	23	37	42	67

For the whole period the Mean Deviation is 39.5.

The Deviations are small, the January Mean Deviation being roughly 3 per cent of the mean production for January; for December it is 3.4 per cent.

On the assumption that the Seasonal Index for any one month is constant for a given time series, it will be seen that Least Squares can be used in several ways to yield Seasonals. I give one example of its use, obtaining Seasonals by a method closely allied to the Link Relative method.

In the Link Relative method link relatives are formed for all the different months. This involves the greater part of the calculation, and it seemed feasible that instead of calculating link relatives and finding median values one could assume that the production for any one month with reference to that for the previous month is given by

February	$= a_1$ January
March	$= a_2$ February
. . . .	. . . .
. . . .	. . . .
December	$= a_{11}$ November
January	$= a_{12}$ December

where, as before, the name of the month stands for the production for that month, and $a_1, a_2, \ldots, a_{12}$ are constants which can be determined by the Method of Least Squares. For February $= a_1$ Janu-are, we have

$$a_1 = \frac{\Sigma\,(\text{February})\,(\text{January})}{\Sigma\,(\text{January})^2}$$

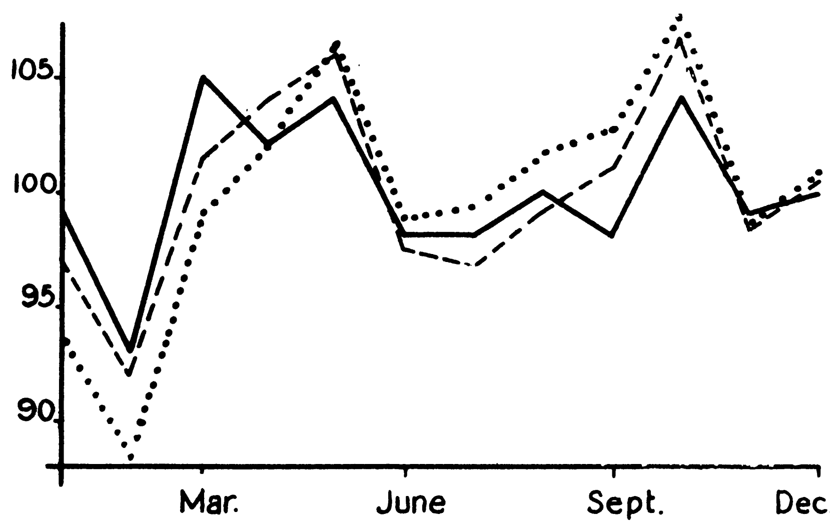
the summations extending over the years of the time series.

Considering our time series to be $u_1, u_2, u_3, \cdots u_{12}, u_{13}, \cdots$

$$a_1 = \frac{u_1 u_2 + u_{13} u_{14} + u_{25} u_{26} + \cdots}{u_1^2 + u_{13}^2 + u_{25}^2}$$

FIGURE I

——— Actual Seasonals.
– – – New Interpolation Method.
..... Detroit Edison Interpolation.



i. e., the observed productions for two successive months are multiplied together and summed and the whole divided by the sum of the squares of the production of the first of the months.

These coefficients a_1, a_2, . . . correspond to the median link relatives, and the procedure is then similar to that used in that method, i. e., January is assumed to be 100.0, etc. We thus get the following table.

TABLE III

Month	(1) a_r	(2) Chain Relative	(3) (2) Adjusted	(4) Seasonal Indices	(5) Error from Actual divided by 100
January	.951	100.0	100.0	95.5	-.035
February	1.106	95.1	94.8	90.5	-.025
March	1.043	105.2	104.3	99.6	-.054
April	1.037	109.6	108.3	103.4	+.014
May	.934	113.7	111.9	106.8	+.028
June	.997	106.2	104.1	99.4	+.014
July	1.018	106.0	103.5	98.8	+.008
August	1.019	107.9	105.0	100.3	+.003
September	1.056	110.0	106.6	101.8	+.038
October	.915	116.2	112.2	107.1	+.031
November	1.020	106.3	102.2	97.6	-.014
December	.967	108.4	103.9	99.2	-.008
January		104.8	100.0		

The Standard Deviation of the Errors of column (5) is found to be ± 0.0269, which is considerably less than that of the Link Relative Method.

If we assume that ${}_{\bullet}u_x$ can be represented by the points on a theoretical curve ${}_{\bullet}u_x = s(x)\ f(x)\ c(x) + \varepsilon_x$ as given by the Detroit Edison Statistical Department, it will be seen that February a_1 (January) gives

$$a_1 = \frac{s(2)}{s(1)}\ \frac{\Sigma[f(x)\ c(x)][f(x+1)\ c(x+)]}{\Sigma[f(x)\ c(x)]^2}$$

where, if $x = 1$ corresponds to the first January of time series, $x = 1$, 13, 25, etc.

$\frac{S(2)}{S(1)}$ can therefore be found as soon as a value can be obtained for the adjustment factor $\frac{\Sigma\, \psi(x)\, \psi(x+1)}{\Sigma\, [\psi(x)]^2}$ where $\psi(x) = f(x)\, c(x)$ If the time series is smoothed by the method already discussed, satisfactory values of $\psi(x)$ are obtained and the adjustment factors easily computed. The smoothed values of the ordinates of the theoretical time series are given in Table II. As logarithmic correction, which has already been employed, assumes a constant adjustment factor for any pair of consecutive months, it will be interesting to find whether the assumption of a theoretical curve for the time series yields better adjustment factors than the constant one used in logarithmic correction.

TABLE IV

Month	Adjustment (1)	Chain Relative (2)	(2) adjusted (3)	Seasonal (4)	Error from actual/100
January		100.0	100.0	97.2	-.018
February	.994	94.5	94.4	91.8	-.012
March	.993	103.8	103.6	100.7	-.043
April	.991	107.3	107.1	104.1	+.021
May	.980	109.0	108.7	105.7	+.017
June	.984	100.2	98.8	96.0	-.020
July	.996	99.5	99.0	96.2	-.018
August	1.000	101.3	100.8	98.0	-.020
September	1.015	104.7	104.1	101.2	+.032
October	1.016	112.3	111.6	108.5	+.045
November	1.006	103.3	102.5	99.6	+.006
December	.995	104.8	104.0	101.1	+.011
January	.996	100.9	100.0		

The adjustment factors are given in column (1) of Table IV, and the corresponding seasonal indices in column (4). The standard deviation of errors, ± .0248, is less than that obtained with the adjustment as used in the link relative, but in this particular case the adjustment factor, using logarithmic correction, would be .996 for each month, differing little from the factors using smoothed ordinates. It will be noted that, owing to accidental errors, the chain relative for January is 100.9, not 100, and an arithmetical correction has to be applied.

From this one sees that, taking accidental errors into account, logarithmic correction is well adapted for reduction purposes.

Finally, I found the Seasonal Indices by the Variate Difference Method. In this method the trend is removed and second differences taken, which are treated by Fourier Analysis. For the second differences I obtained

$$\Delta^2 u = +0.02 + 0.822 \cos(\theta - 339°52') + 3.958 \cos(2\theta - 314°53') + 3.902 \cos(3\theta - 39°34') + 4.374 \cos(4\theta - 25°58') + 9.942 \cos(5\theta - 293°39') + 0.775 \cos 6\theta.$$

yielding the seasonal indices:

Jan.	96.9	May	105.6	Sept.	101.9
Feb.	89.2	June	99.4	Oct.	106.3
Mar.	100.8	July	98.3	Nov.	97.8
Apr.	103.4	Aug.	102.2	Dec.	98.3

Dividing the seasonals by 100 and comparing with Actual values, Mean and Standard Deviation of errors from the actual seasonals are

TABLE V

SEASONAL INDICES

Month	Actual Values	Interpolation: Detroit Edison	Interpolation: Robb	Link Relative	Modified Link Relative: Log. Corec-tion	Modified Link Relative: Theoretical Correction	Variate Differ-ence
January	.990	.938	.973	.975	.955	.972	.969
February	.930	.885	.919	.890	.905	.918	.892
March	1.050	.988	1.014	.988	.996	1.007	1.008
April	1.020	1.021	1.039	1.007	1.034	1.041	1.034
May	1.040	1.065	1.062	1.030	1.068	1.057	1.056
June	.980	.986	.974	.962	.994	.960	.994
July	.980	.993	.967	.972	.988	.962	.983
August	1.000	1.017	.993	1.013	1.003	.980	1.022
September	.980	1.028	1.010	1.033	1.018	1.012	1.019
October	1.040	1.079	1.067	1.099	1.071	1.085	1.063
November	.990	.985	.982	1.004	.976	.996	.978
December	1.000	1.009	1.005	1.027	.992	1.011	.983
S. D.		.0337	.0194	.0338	.0269	.0248	.0246

given by ± 0.022 and ± 0.0246, which are still considerably less than those obtained by Link Relative or Detroit Edison Interpolation. For reference I have put in Table V the results obtained from all the methods mentiond in this paper, together with their Standard Deviation (S. D.) of errors.

As the time taken to determine the Seasonal Indices by the various methods is important, I took the time series of Merchandise Imports for a period of ten years, and calculated the Seasonal Indices. Denoting the Link Relative method by R_1, Modified Link Relative (Log. correction) by R_2, Interpolation (as given in this paper) by I, I found that as regards the time taken for one determination completely checked $R_1 : R_2 : I :: 11 : 8 : 7$

For the particular example taken, it took 1.75 hours to transcribe the material and to determine the Indices, completely checked, by the Interpolation method. For the Link Relative method, 7.75 hours were taken. It is desirable, if possible, to have two independent determinations, and the above times would consequently have to be doubled. The Variate Difference method roughly takes the same time as the Link Relative method, when the trend has been removed from the time series.

Richard A. Robb.

Glasgow, Scotland.