Direct estimation of seasonal variation.

by

C. E. V. Leser

The Economic Research Institute
Dublin.

September 1965.

Direct estimation of seasonal variation.

In estimating the seasonal component of a time series which is assumed to contain a trend and a residual in addition, the usual procedure is to specify, estimate and eliminate the trend before proceeding to the estimation of seasonal variation. Modifications are possible which permit the simultaneous estimation of trend and seasonal variation, but the results still depend on the assumptions made with regard to the form of trend.

The three components are taken here to be additive, if necessary after a logarithmic transformation of the data. In these circumstances, the assumption of a linear trend has some attractive features, as it theoretically permits the application of significance tests and the construction of confidence intervals for the seasonal components as well as the trend slope. In practice, there are often difficulties, as with extended time series there are usually marked departures from linearity present in the trend. If then the trend is taken as linear, this means that the residuals include a substantial cyclical element and are not random errors but autocorrelated. Thus the estimated residual variance will be excessively large, and at the same time the underlying assumptions for the application of standard error formulae are not satisfied.

However, the validity of the formulae for the seasonal component estimation which result from the assumption of a linear trend does not depend on this assumption. It will be shown here that much weaker assumptions suffice to justify this method of estimating the seasonal variation. The discussion will be given for quarterly data but may immediately be extended to monthly or other data.

Given a period of m calendar years, write Y_{ij} for a variable observed in the jth quarter of the ith year,

T_{ij} for the trend, S_j for the seasonal variation and E_{ij} for the residual; such that

$$Y_{ij} = T_{ij} + S_j + E_{ij} \quad (i = 1, \ldots, m; \quad j = 1, 2, 3, 4) \qquad (1)$$

$$\sum_{j=1}^{4} S_j = 0 \qquad (2)$$

The following assumptions will now be made:

(a) For each quarter, the m residuals add up to zero.

(b) The annual totals of residuals are orthogonal to time as measured in years.

(c) The four quarterly totals of trend values form an arithmetic progression; and their successive differences are in the same proportion to the regression coefficient of annual trend totals on time in years as if the trend was fully linear.

The first two assumptions are easily formulated as follows:

$$\sum_{i=1}^{m} E_{ij} = 0 \qquad (j = 1, 2, 3, 4) \qquad (3)$$

$$\sum_{i=1}^{m} i \sum_{j=1}^{4} E_{ij} = 0 \qquad (4)$$

To formulate the third assumption, write

$$z_i = 2i - m - 1 \qquad (i = 1, \ldots, m)$$

so that z_i goes from $-(m-1)$ in steps of 2 to $m-1$, and $\sum_{i=1}^{m} z_i = 0$. Then the regression coefficient β of annual trend totals $\sum_{j=1}^{4} T_{ij}$ on z_i is

$$\beta = 3 \sum_{i=1}^{m} z_i \sum_{j=1}^{4} T_{ij}/(m^3 - m)$$

Now if the trend is linear, with differences D between successive quarterly terms, the difference between successive quarterly totals is mD, and the regression coefficient of annual totals on z is 8D. The third assumption thus is

$$\sum_{i=1}^{m} (T_{i\ j+1} - T_{ij}) = m\beta/8 \qquad (j = 1, 2, 3)$$

$$\left.\begin{aligned} \text{or: } \sum_{i=1}^{m} (T_{i\ j+1} - T_{ij}) &= \lambda \qquad (j = 1, 2, 3) \\ \lambda &= 3 \sum_{i=1}^{m} z_i \sum_{j=1}^{4} T_{ij}/8(m^2-1) \end{aligned}\right\} \qquad (5)$$

Now write A_i for the annual and Q_j for the quarterly totals of the observations, thus

$$A_i = \sum_{j=1}^{4} Y_{ij} \qquad (i = 1, \ldots, m)$$

$$Q_j = \sum_{i=1}^{m} Y_{ij} \qquad (j = 1, 2, 3, 4)$$

From (1) and (3)

$$Q_j = \sum_{i=1}^{m} T_{ij} + mS_j \quad (j = 1, 2, 3, 4) \qquad (6)$$

From (1) and (2)

$$A_i = \sum_{j=1}^{4} T_{ij} + \sum_{j=1}^{4} E_{ij} \quad (i = 1, \ldots, m)$$

and with the help of (4) we obtain

$$\sum_{i=1}^{m} z_i A_i = \sum_{i=1}^{m} z_i \sum_{j=1}^{4} T_{ij} \qquad (7)$$

(5), (6) and (7) combined yield the three equations

$$m(S_j - S_{j+1}) = (Q_j - Q_{j+1}) + 3 \sum_{i=1}^{m} z_i A_i/8(m^2 - 1) \qquad (8)$$

$$(j = 1, 2, 3)$$

which together with (2) determine the values of S_1, S_2, S_3 and S_4. Writing for the grand total

$$G = \sum_{i=1}^{m} \sum_{j=1}^{4} Y_{ij} = \sum_{i=1}^{m} A_i = \sum_{j=1}^{4} Q_j$$

the explicit solution is

$$S_j = \{4(m^2 - 1)(4Q_j - G) + 3(5 - 2j) \sum_{i=1}^{m} z_i A_i\} / 16m(m^2-1) \qquad (9)$$

$$(j = 1, 2, 3, 4)$$

It can easily be shown that (9) is the solution obtained by assuming a fully linear trend and by fitting it with the help of least squares. Write

$$x_{ij} = 2(4i + j - 4) - 4m - 1$$
$$= 4 z_i + (2j - 5)$$

so that $x_{11} = -4m+1$, $x_{12} = -4m+3$, ..., $x_{m4} = 4m-1$. Then the condition

$$\sum_{i=1}^{m} \sum_{j=1}^{4} (Y_{ij} - a - S_j - b\, x_{ij})^2 = \text{Minimum}$$

subject to the constraint (2) leads to the following normal equations

$$4\, m\, a = \sum_{i=1}^{m} \sum_{j=1}^{4} Y_{ij}$$

$$m\, a + m\, S_j + b \sum_{i=1}^{m} x_{ij} = \sum_{i=1}^{m} Y_{ij} \qquad (j = 1,2,3,4)$$

$$\sum_{j=1}^{4} S_j \sum_{i=1}^{m} x_{ij} + b \sum_{i=1}^{m} \sum_{j=1}^{4} x_{ij}^2 = \sum_{i=1}^{m} \sum_{j=1}^{4} x_{ij} Y_{ij}$$

Using the notation adopted above, and noting that

$$\sum_{i=1}^{m} x_{ij} = (2j - 5)\, m \qquad (j = 1,2,3,4)$$

$$\sum_{i=1}^{m} \sum_{j=1}^{4} x_{ij}^2 = 4\, m\, (16m^2 - 1)/3$$

we have

$$a = G/4\, m = \dot{\bar{Y}}$$

and the remaining equations are

$$\left.\begin{aligned} 4\, m\, \{S_j + (2j - 5)\, b\} &= 4\, Q_j - G \qquad (j = 1,2,3,4) \\ 3\, m \sum_{j=1}^{4} (2j - 5)\, S_j + 4m(16m^2 - 1)b &= 12 \sum_{i=1}^{m} z_i A_i + 3 \sum_{j=1}^{4} (2j-5) Q_j \end{aligned}\right\} \qquad (10)$$

The solution for S_1, S_2, S_3 and S_4 given by (9) together with the solution for b

$$b = 3 \sum_{i=1}^{m} z_i \; A_i / 16m \; (m^2 - 1) \qquad (11)$$

satisfy the five equations (10), q.e.d.

Formula (9) permits an easy computation of the seasonal variation estimates, with the help of the quarterly totals and a linear combination of the annual totals. For specific values of m not divisible by 6, the formula may be further simplified, since with m odd, all values of z_i are divisible by 2, and if m is not a multiple of 3, $m^2 - 1$ is divisible by 3. The formula is applicable even in the case m = 2.

(9) gives the seasonal variation as an additive component. If the seasonal components are believed to be proportionate to the trend and it is not desired, for theoretical or practical reasons, to use a logarithmic transformation, the seasonal components may be expressed as percentages of the overall mean $\overline{Y}$, to obtain a set of seasonal indices with mean 100 by addition to or subtraction from 100.

It may sometimes be of interest to ascertain which portion of the variation between quarters within calendar years is accounted for by the seasonal component, and possibly which proportion by the average trend, the exact trend being unknown. In the present terminology, we have

$$\text{Total S Sq} = \sum_{i=1}^{m} \sum_{j=1}^{4} (Y_{ij} - A_i/4)^2$$

and it is easily seen that

$$\text{S sq. seas. var.} = 2 \sum_{j=1}^{4} S_j \, Q_j - m \sum_{j=1}^{4} S_j^{\,2} \qquad (12)$$

$$\text{S sq. av. trend} = 20 \, m \, b^2 \qquad (13)$$

where b is given by formula (11).

Example: Value of imports, Ireland 1960-64 (£Mill)

i \ j (Y_{ij})	1	2	3	4	A_i
1	57.0	55.9	52.2	61.2	226.3
2	65.8	67.4	62.3	65.8	261.3
3	67.3	67.3	64.7	74.3	273.6
4	69.4	80.0	69.9	87.6	306.9
5	87.7	91.1	81.2	87.9	347.9
Q_j	347.2	361.7	330.3	376.8	1,416.0

$$m = 5, \quad m^2 - 1 = 24$$

$$z_1 = -4, \; z_2 = -2, \; z_3 = 0, \; z_4 = 2, \; z_5 = 4$$

$$S_1 = \{16\,(4Q_1 - G) + 3\,(2A_5 + A_4 - A_2 - 2A_1)\}/320$$

$$S_2 = \{16\,(4Q_2 - G) + (2A_5 + A_4 - A_2 - 2A_1)\}/320$$

$$S_3 = \{16\,(4Q_3 - G) - (2A_5 + A_4 - A_2 - 2A_1)\}/320$$

$$S_4 = \{16\,(4Q_4 - G) - 3\,(2A_5 + A_4 - A_2 - 2A_1)\}/320$$

$$2A_5 + A_4 - A_2 - 2A_1 = 288.8$$

$S_1 = 1.4$, $S_2 = 2.4$, $S_3 = -5.6$, $S_4 = 1.8$

or since $\overline{Y} = 70.8$, we have for seasonal indices S_j':

$S_1' = 102.0$, $S_2' = 103.5$, $S_3' = 92.0$, $S_4' = 102.5$

Furthermore

$$\sum_{i=1}^{5} \sum_{j=1}^{4} (Y_{ij} - A_i/4)^2 = 102,779.06 - 102,392.09 = 386.97$$

$$2 \sum_{j=1}^{4} S_j Q_j - m \sum_{j=1}^{4} S_j^2 = 2 \times 182.72 - 5 \times 42,32 = 153.84$$

$$b = 288.8/320 = .9025$$

$$20m\, b^2 = 81.45$$

Thus the seasonal variation explains about 40% and the average trend increase about 21% of the variation between quarters within a calendar year over the five-year period, the remaining 39% being accounted for by deviations from linearity in the trend and by irregular fluctuations.

The method discussed here is particularly useful when no trend estimate is required and it is only a question of obtaining seasonally corrected data. It is, of course, still possible to specify a trend which satisfies the assumptions already made and to estimate it from the seasonally corrected data.

The method may be applied to two or more time series subject to a linear relationship which is to be estimated. If a linear trend as well as the seasonal component are eliminated from each time series, then the result is the same as introducing a linear trend and dummy variables for the seasons into the regression. This procedure, however, assumes a linear trend to be appropriate which is in many instances unrealistic.

The elimination of seasonal variation alone and the application of least squares to the seasonally corrected data does not yield the same result as the application of least squares to the original data with the help of dummy variables, which is theoretically preferable on grounds of statistical efficiency. However, seasonal correction of individual series has the practical advantage of considerably simplifying the computations; this is particularly valuable when various alternative regressions are being investigated.