COWLES COMMISSION DISCUSSION PAPERS: STATISTICS, No. 325

On the decomposition of a series of observations composed of a trend, a periodic movement with known period and a stochastic variable by means of a moving average.

by A. Hald

January 14, 1949

The following remarks contain a summary of the main results in § 11 of my paper on "The decomposition of a series of observations," Nordisk Tidskrift for teknisk Økonomi, 1947, and further a few supplementary considerations. References are to that paper.

Notations:

$M\{y\}$ = mean value (expectation) of y.

$V\{y\}$ = variance of y.

$V\{y_1, y_2\}$ = covariance of (y_1, y_2).

$\bar{y}_{\nu i}$ = a moving average of y's with $y_{\nu i}$ as central element.

$y_{.i}$ = a mean of $y_{\nu i}$'s with regard to ν .

$\bar{y}_{.i}$ = a mean of $\bar{y}_{\nu i}$'s with regard to ν .

The specification.

The observations consist of nk pairs of values $(x_{\nu i}, y_{\nu i})$, ν =1,...., n and i = 1, ..., k, where the values of the independent (fixed) variable x are assumed to be equidistant and the values of the dependent variable are assumed to be composed of three terms

$$y_{\nu i} = \Gamma(x_{\nu i}) + \eta_i + \varepsilon_{\nu i} \quad ,$$

$\Gamma(x)$ denoting the trend, η the periodic movement and ε the stochastic component, cf. Fig. 4.1.

In an economic time series x denotes time, y the observed economic variable, n the number of years, k the number of observations within a year, $\Gamma(x)$ the trend, $\eta_1, \ldots, \eta_k$ the values of the seasonal component and ε the random variable, cf. Fig. 18.1.

In an agricultural field experiment with the plots systematically arranged in a row x denotes the position (number) of the plot, y the observed yield, n the number of blocks, k the number of treatments, $\Gamma(x)$ the effect of systematic factors (fertility), which we wish to eliminate, η the effect of the treatments, and ε the effect of the experimental error, cf. Fig. 15.1.

The properties of the three components of y are usually specified in the following way:

1) The ε's are independent and normally distributed with mean 0 and variance σ^2.

2) No constraints are imposed on the η's.

3) The functional form of $\Gamma(x)$ is chosen in a suitable manner. In my paper the problem of decomposition is treated with three different specifications of $\Gamma(x)$:

a) $\Gamma(x_{\nu i}) = \Gamma_\nu$ for i = 1, ..., k, i.e. $\Gamma(x)$ is a step function. The corresponding statistical technique is the analysis of variance, cf. § 10.

b) $\Gamma(x)$ is chosen as a polynomial in x. The statistical technique is then the regression analysis, cf. §§ 4-7. For practical applications it pays to work out a special technique based on generalised orthogonal polynomials, cf. §§ 8-9.

c) $\Gamma(x)$ is "approximately linear" in every interval of length k. The corresponding statistical technique is the method of moving averages, cf. § 11.

From a purely mathematical point of view the method of moving averages assumes that the functional form of $\Gamma(x)$ is such that to satisfy the difference equation

$$\Gamma(x-m) + \dots + \Gamma(x) + \dots + \Gamma(x+m) = (2m+1)\,\Gamma(x) ,$$

where $k = 2m + 1$. The solution of this difference equation has been given by W. Simonsen in Nordisk Tidsskrift for teknisk Økonomi, 1948. (A comparison of the method of moving averages in this case and the regression analysis using this functional form is not yet completed).

The model can obviously be generalised in different ways, e.g. by assuming that neighboring $\mathcal{E}$'s are correlated.

The model is purely descriptive; no attempt is made to "explain" the formation of the observations.

In cases where the assumption about an additive composition does not hold good it may sometimes be replaced by the assumption that the composition is multiplicative, which means that the logarithms of the y's can be treated in accordance with the above specification.

The method of moving averages.

Only the case where k is odd, $k = 2m + 1$, will be treated here, cf. pp. 36-57. The case where k is even is in principle analogous to k odd but leads to more complicated algebra, cf. pp. 57-64.

The specification is

$$y_{\nu i} = \Gamma_{\nu i} + \eta_i + \varepsilon_{\nu i} ,$$

where $\Gamma_{..} = 0$ and $\eta_. = \eta$.

The moving average of extent k "eliminates" the periodic component as

$$\bar{y}_{\nu i} = \bar{\Gamma}_{\nu i} + \eta + \bar{\varepsilon}_{\nu i} .$$

It is assumed that the difference $\bar{\Gamma}_{\nu i} - \Gamma_{\nu i}$ is negligible as compared with the random variation, so that we can ignore the systematic error $\bar{\Gamma}_{\nu i} - \Gamma_{\nu i}$, i.e. we get

$$\bar{y}_{\nu i} = \Gamma_{\nu i} + \eta + \bar{\varepsilon}_{\nu i} .$$

By means of the moving averages we eliminate the trend from the series of observations:

$$d_{\nu i} = y_{\nu i} - \bar{y}_{\nu i} = \eta_i - \eta + \varepsilon_{\nu i} - \bar{\varepsilon}_{\nu i} .$$

We then get the following unbiased estimates:

$$\hat{\Gamma}_{\nu i} = \bar{y}_{\nu i} - y_{..} ,$$

$$\hat{\eta}_i = y_{.i} - \bar{y}_{.i} + y_{..} ,$$

$$Y_{\nu i} = \hat{\Gamma}_{\nu i} + \hat{\eta}_i = \bar{y}_{\nu i} + y_{.i} - \bar{y}_{.i} ,$$

$$s_1^2 = \frac{1}{(n-1)(k-1)} \sum_{\nu=1}^{n} \sum_{i=1}^{k} (y_{\nu i} - Y_{\nu i})^2 .$$

The distribution of $(\hat{\eta}_1, \ldots, \hat{\eta}_k)$ is a k-dimensional normal distribution with moment matrix given by formula (11.61).

The variance of s_1^2 is given by (11.107).

Tests of significance analogous with the tests known from the analysis of variance are discussed on pp. 56-57.

Applications.

Nearly all agricultural field experiments are based on models additively composed of three elements of the above mentioned type. The question in designing an experiment is whether to use randomization and analysis of variance or systematic arrangements and regression analysis or moving averages, cf. § 16. (The best method is to use randomization and regression analysis but this method leads to very laborious computations).

The applicability of the model in the description of economic time series is without the range of my experience, as I have only treated the case mentioned in § 18.