On Rate of Change Per Cent Per Annum Over a Period of Years.

The problem arose in considering a recent draft paper in ESRI. Let the observations be $Y_1, Y_2, \ldots, Y_T$ over a period of years T. The intuitive method is to set -

(1) $(I + T)^{T-1} = Y_T/Y_1,$

where 100 r is the rate % per annum, calculated from logarithms -

(2) $\log(1 + r) = (\log Y_t - \log Y_1)/(T - 1).$

An immediate objection is that the calculation relies solely on the first and last observations (ignoring the remaining (T - 2) observations) and either or both of these might be manifestly abnormal in relation to adjacent observations. Yet the calculation may find a kind of justification in the theory of averages, for

(3) $Y_T/Y_1 = (Y_2/Y_1)\ (Y_3/Y_2)\ \ldots\ (Y_{T-1}/Y_{T-2})\ (Y_T/Y_{T-1}),$

the right side apparently taking all the individual changes into account, (1 + r) being the geometric mean of the series. The answer is, of course, that one cannot say one has taken account of, say, T_2 when having brought it in one proceeds to take it out.

The better way (as we shall see) is to fit an exponential curve to the data:-

(4) $Y_t = Ce^{\beta t} + e_t,$

the coefficient β, the annual rate of change to be estimated by LS regression from

(5) $\log_{10} Y_t = \alpha + \beta t \log_{10} e + u_t$

or, with obvious change of notation,

(6) $y_t = \alpha + \beta' t,$

so that, if b and b' are respectively the regression estimates of β and β', -

(7) $b' = b \log_{10} e = \Sigma(y_t - \bar{y})\ (t - \bar{t})/\Sigma(t - \bar{t})^2.$

We shall now compare the relative efficiency of the two methods of estimation on stochastic lines. Assume that the economic series (in log form) over a period of years can be represented by -

(8) $Y_t = \alpha + \beta t + u_t$, $t = 1, 2, \ldots, T$

the residual u_t being regular (i.e. mean zero, homoskedastic, variance σ^2, elements mutually uncorrelated). This assumption is approximately valid (at least enough so for the present purpose) for most Irish economic series in the postwar period. The first estimate of β, namely b_1, is given by

(9) $b_1 = (Y_T - Y_1)/(T - 1) = \beta + (u_T - u_1)/(T - 1)$

Since $E(b_1) = \beta$, b_1 is an unbiased estimate of β. Its variance is -

(10) $\text{Var } b_1 = E(b_1 - \beta)^2 = 2\sigma^2/(T - 1)^2$

The second estimate of β, i.e. the regression estimate b_2 is

(11) $b_2 = \Sigma(Y_t - \bar{Y})(t - \bar{t})/\Sigma(t - \bar{t})^2$,

so that, as is well-known, b_2 is an unbiased estimate of β and, from (11),

(12) $\text{Var } b_2 = E(b_2 - \beta)^2 = \sigma^2/\Sigma(t - \bar{t})^2$.

Now

(13) $\Sigma(t - \bar{t})^2 = T(T^2 - 1)/12$.

Hence

(14) $\text{Var } b_2 = 12\sigma^2/T(T^2 - 1)$.

Comparing var b_1 and var b_2 from (10) and (14) the two methods are equally efficient for $T = 2$ (obvious *a priori* and therefore checking the algebra) and (more curiously) $T = 3$. For $T > 3$ the relative efficiency of b_2 increases rapidly, in fact as $O(T)$. For $T = 10$, for example, var b_1/var b_2 = 2.04 and for $T = 20$ to 3.68. For large values of T the relative efficiency is very nearly $T/6$.

1 March 1970.

R. C. Geary