Two Exercises in Simple Regression

R. C. GEARY

WHILE the two exercises are independent in that they purport to answer separate questions, methodologically they are very similar. They both compare the efficiency of different methods of estimation of parameters by the familiar method of ratio F of variances of estimate, stochastically decisive, under very general conditions, when number of observations T tends towards infinity. It happens that variance formulae in the two parts are formally identical, namely var b (1.13) and var b_2 (2.10); also var b' (1.14) and var b_1 (2.6), for reasons which, no doubt, will be evident.

1. SHOULD REGRESSION IN TIME SERIES BE COMPUTED USING ORIGINAL DATA OR THEIR DELTAS?

All practitioners are aware that least squares regression equations in economic time series are likely to produce r's very near unity because during the regression period (e.g. post-war) all series tend to have the same trend. Such high correlations can never be regarded *ex post* as indicating relationship except when there are good theoretical reasons for suspecting relationship, e.g. between income and consumption. To establish short term relationship two practices are common:

(ii) Introduce time t as an indvar.

(ii) Instead of using the original variables (in simple regression say Y_t and X_t) use ΔY_t and ΔX_t, with $\Delta Y_t = Y_t - Y_{t-1}$.

Both procedures have the effect of eliminating trend t. With (i) indvars would be X_t and t. The coefficient of X_t would be *exactly* that which would be found between the *residuals* of Y_t and X_t after regression on t.

In what follows we deal only with (ii). It will suffice to confine attention to simple (i.e. two-variable) regression.

Models

Let the model be—

$$Y_t = \alpha + \beta X_t + u_t, \; t = 1, 2, \ldots, T. \tag{1.1}$$

where the error term u_t is *regular* (i.e., for all t, $Eu_t = 0$, $Eu_t^2 = \sigma^2$. $Eu_{t'}u_t$ $(t' \neq t) = 0$). Then, if b be the regression estimate of β, with $x_t = X_t - \bar{X}$,—

$$b = \beta + \Sigma x_t u_t / \Sigma x^2_t \tag{1.2}$$

so that $Eb = \beta$ with variance—

$$\text{var } b = E(b-\beta)^2 = \sigma^2 / \Sigma x_t^2, \tag{1.3}$$

classical results, of course. The Δ version of (1.1) is—

$$Y'_t = \beta X'_t + u'_t, \; t = 2, 3, \ldots, T, \tag{1.4}$$

where $Y'_t = \Delta Y_t = Y_t - Y_{t-1}$, etc. There are now $(T-1)$ terms. The error term u'_t is, however, no longer regular since obviously $Eu'_t u'_{t-1} = -\sigma^2$, not zero. Also $Eu'^2_t = 2\sigma^2$.

Maximum Likelihood

If (1.4) (with an additional relation from (1.1), say $Y_1 = \alpha + \beta X_1 + u_1$, to make T relations in all) are both solved by maximum likelihood (ML) the estimates from a given realisation of α and β will be identical. In fact, if the probability element of the vector u is—

$$f(u_1, u_2, \ldots, u_T) \Pi du_t \tag{1.5}$$

the ML solution is found as the values of the parameters (e.g. α, β) which maximise f, regarding the u_t in f as functions of the parameters and the data (e.g. as given by (1.1). If in (1.5) we make the linear transformation (in matrix form)—

$$\mathbf{u}' = \mathbf{A}\mathbf{u}, \tag{1.6}$$

where prime on left is *not* transpose and $\mathbf{A}$ is any non-singular square matrix with pre-determined numerical elements, (1.5) transforms into—

$$| \mathrm{A}^{-1} | \; g(u'_1, u'_2, \ldots, u'_T) \Pi du'_t, \tag{1.7}$$

where g is the function f after the transformation. Since in (1.7) the absolute value of the determinant, namely $| \mathrm{A}^{-1} |$, is a constant (i.e. independent of the parameters), the problem of the maximisation of f and g are identical and the maximising values of the parameters the same. In our particular application the matrix **A** is—

$$\mathbf{A}=\begin{pmatrix} 1 & 0 & 0 \dots & 0 & 0 \\ -1 & 1 & 0 \dots & 0 & 0 \\ 0 & -1 & 1 \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 \dots & -1 & 1 \end{pmatrix} \tag{1.8}$$

LS Regression Applied to Delta Version (1.4)

Invariably, however, when the regression problem is deltaized the assumption is made that the error term u'_t is regular, which assumption amounts to a wrong specification *if the basic model is* (1.1). Usually a constant term is added, which would indeed by formally consistent with the model—

$$Y_t = \alpha+\beta X_t+\gamma t+u_t, \tag{1.9}$$

i.e. as at (i) in the opening paragraph.

When (1.4) is regarded as a problem in least squares the estimate b' of β is—

$$\begin{aligned} b' &= \Sigma X'_t Y'_t / \Sigma X'^2_t \\ &= \beta+\Sigma(x_t-x_{t-1})(u_t-u_{t-1})/\Sigma(x_t-x_{t-1})^2 \end{aligned} \tag{1.10}$$

so that b' is still an unbiased estimate of β, i.e., $Eb' = \beta$.

Its variance—

$$\operatorname{var} b' = 2\sigma^2(\Sigma x'^2_t - \Sigma x'_t x'_{t-1})/(\underset{1}{\Sigma x'^2_t}),^2 \tag{1.11}$$

recalling that $X'_t = X_t - X_{t-1} = x_t - x_{t-1} = x'_t$ and that now the Σ's on the right have $(T-1)$ or $(T-2)$ terms.

We apply the foregoing theory first to two particular cases, then asymptotically to a more general case.

Case 1

Let $X_t = t(t = 1, 2, \dots, T)$, the very common equi-spaced indvar case. Then—

$$\Sigma x^2_t = T(T^2-1)/12 \tag{1.12}$$

and, from (1.3),—

$$\operatorname{var} b = 12\sigma^2/T(T^2-1). \tag{1.13}$$

All the x'_t are unity, so that, from (1.11),—

$$\text{var } b' = 2\sigma^2/(T-1)^2, \tag{1.14}$$

so that, if the efficiency F of b' in relation to b = var b/var b',—

$$F = 6(T-1)/T(T+1) \tag{1.15}$$

The methods are equally efficient ($E = 1$) for $T = 2, 3$. Thereafter the efficiency of b' diminishes rapidly, in fact approximately as $6/T$.

Case 2

Often we notice a tendency for the indvar values to cluster near the median so that our second constructed example will illustrate this. Let there be $2T$ observations so that X_t is—

$$-T^2, -(T-1)^2, \ldots, -2^2, -1^2, 1^2, 2^2, \ldots, (T-1)^2, T^2.$$

Using the sum S_4 of the fourth powers of the natural numbers $1, 2, \ldots, T$, namely

$$S_4 = T(T+1)(2T+1)(3T^2+3T-1)/30, \tag{1.16}$$

we find from (1.3)—

$$\text{var } b = 15\sigma^2/T(T+1)(2T+1)(3T^2+3T-1) \tag{1.17}$$

having noted that $\bar{X} = 0$ so that $X_t = x_t$.

As regards var b', the sequence x'_t is—

$$2T-1, 2T-3, \ldots, 5, 3; 2, 3, 5, \ldots, 2T-3, 2T-1,$$

$(2T-1)$ terms in all, so that, after some elementary algebra, and using (1.11),—

$$\text{var } b' = 18\sigma^2(4T^2-6)/(8T^3-2T+6)^2 \tag{1.18}$$

and efficiency F of estimate b' is—

$$F = 5(8T^3-2T+6)^2/6T(T+1)(2T+1)(3T^2+3T-1)(4T^2-6), \tag{1.19}$$

tending to $5/18T$, when T is large. We recall, however, that for this Case 2, number of observations is not T, but $2T = T'$, say, whence limiting value of F is $20/9T'$, in comparison with $6/T'$ for Case 1. In the more typical Case 2 the efficiency of the delta procedure estimate b' (in relation to b) is even worse than in Case 1. Both are very bad.

A Remark

The variances of b given by (1.13) and (1.17) are not $O(T^{-1})$ as in classical theory. In fact, in the foregoing *exposé* no regard was paid to orders of magnitude. This would have been achieved by multiplying all the indvar values as given by KT^{-1} in Case 1 and by KT^{-2} in Case 2, where K is independent of T. This treatment would render both estimates of var b' of order $O(T^0)$, an ordinary magnitude. Using LS with the deltas the estimates b' are no longer consistent, as not tending to β as T tends to infinity. The values of F at (1.15) and (1.19) would not be affected.

The General Polynomial Case

In Cases 1 and 2 X_t was represented by polynomials of degree 1 and 2 in t. Suppose, more generally, that X_t is a polynomial of degree k in t. Then ΣX_t is $O(T^{k+1})$. Using this formula on (1.3), we find var b is $O(T^{-2k-1})$. With reference to (1.11), $\Delta X_t = x'_t$ is a polynomial of degree $(k-1)$ in t. The denominator is $[O(T^{2k-1})]^2 = O(T^{4k-2})$. As to the numerator, the term $(x'^2_t - x'_t x'_{t-1})$ is a polynomial of degree $(2k-3)$ in t, so that its Σ is $O(T^{2k-2})$. Hence var b' is $O(T^{2k-2})/O T^{4k-2}) = O(T^{-2k})$. Then $F =$ var b/var $b' = O(T^{-1})$ as in Cases 1 and 2. If dimensional multipliers of the indvars KT^{-k} be introduced to render var $b = O(T^{-1})$ as it should, var b' is $O(T^0)$, an ordinary magnitude so that, as T tends to infinity b' does not tend in probability to β.

Consequence of Assumption of u'_t *Regular*

Suppose, on the contrary, that in (1.4) u'_t by the Durbin-Watson d or Geary τ tests, can be regarded as regular, variance σ^2, what are the implications for T_t? Clearly—

$$Y_t = \alpha + \beta X_t + u_t \tag{1.20}$$

but u_t can no longer be regarded as regular, since it is heteroskedastic. In fact—

$$u_t = \sum_{t'=1}^{t} u'_{t'} \tag{1.21}$$

so that var $u_t = t\sigma^2$. Such a situation would be unusual, though error variance, in practice, tends to increase with indvar value. More unusual still would be the Durbin-Watson d approaching the small value of approximately $2/T$ because of the high degree of autocorrelation between the residues u_t.* Let us, nevertheless,

*It must be confessed that violent contrasts between values of d using original and delta versions in simple regression have not been encountered in practice. Perhaps this point is worthy of further examination.

follow it through, wrongly assuming that β can be estimated by b'', using LS regression. Then

$$b'' = \Sigma Y_t(X_t - \overline{X})/\Sigma(X_t - \overline{X}^2) \quad (1.22)$$

which, on substitution from (1.20), becomes

$$b'' = \beta + \Sigma u_t x_t / \Sigma x^2_t, \quad (1.23)$$

so that $Eb'' = \beta$. However—

$$\text{var } b'' = \sigma^2(\sum_{t=1}^{T} tx_t^2 + 2\sum_{t=1}^{T} tx_t \sum_{t'=t+1}^{T} x_{t'})/(\Sigma x^2_t)^2 \geqslant \sigma^2 \quad (1.24)$$

which *seems* to be an ordinary magnitude, i.e. (OT^0): we have little interest in establishing this firmly. If this is true than b' does not tend in probability towards β even when T tends towards infinity.

Even if we satisfy ourselves as to the homoskedacity and non-autoregression of residuals in ΔX_t and ΔY_t regression, i.e. that u_t is regular, we should realise the oddity of these assumptions for the relationship between the original data X_t Y_t. At present decision whether to use the original data or their deltas seems largely a matter of whim or instinct, which is not good enough. *Both* cannot be right and criteria should be used in making a choice.

An Example

With X_t gross national expenditure and Y_t money (annual average) 1949–1965 ($T = 17$) the regression coefficient b for Y_t on X_t is 0·2640 with ESE (estimated standard error) 0·00552, $r = ·997$, while for the deltas $b' = 0·1891$ with ESE = 0·0730, $r = ·81$ (15d.f., $P \leqslant ·001$). The efficiency of b' as an estimate of the theoretical β is only .0057. b is incomparably better than b' as an estimate of β.

On the other hand if one's objective is the estimation of ΔY_t from ΔX_t (perhaps for forecasting) it is better to use the regression with b' than that from Y_t on X_t, deriving the calculated ΔY_{tc} *ex post* from the Y_{tc}. On the same data as above the values of the criterion of goodness-of-fit $\Sigma(\Delta Y_t - \Delta Y_{tc})^2$ are 345 and 448 respectively, so that the delta regression, despite its inferior estimate of b' yields a substantially better calculated value of ΔY_t.

But the efficient estimation of the increments ΔY_t is in conflict with the efficient estimation of Y_t. We mean that

$$\sum_{t'=1}^{t} \Delta Y_{t'c},$$

where $\Delta Y_t{}'_c$ have to be estimated by delta regression is less efficient as an estimate of Y_{tc} than is the value calculated from the direct Y_t on X_t regression by the residual sum squares $\Sigma(Y_t - Y_{tc})^2$ test. This is obvious since the direct Y_t regression by definition minimises this expression.

Conclusion

Estimation of coefficients by LS regression from delta regression will usually be highly inefficient. The hypothesis of residual regularity in the delta form of model is bizarre for its implication with regard to the error term of relation between absolute values. If, however this regularity can be regarded as tenable the delta regression can be used efficiently only for estimating the increments ΔY_t and *not* the Y_t themselves.

There is no reason why these conclusions, based on simple regression, should not apply to multivariate regression or to models of several equations.

Our professional consciences may be uneasy about those very high correlations in the original data. It is certainly consoling to find a satisfactory correlation between the deltas of the data since thereby we can be reasonably sure that the original high correlation was not due solely to the fact that each was closely related to time trend t. This is a role for the deltas. Better still to regress on X_t and t together and to find a significant coefficient for X_t. If t is also significant (and the residual non-autoregressed) we have a reasonable forecasting equation.

A point to assuage our tortured consciences. If the indvars we know are all strongly correlated with time trend t, it is plausible to assume that those we *don't* know have the same property. The indvar t may in a certain measure act as a proxy for these, instead of requiring the error term to carry all the brunt. Time trend t may be a more respectable indvar than we customarily think. If t has a significant coefficient residual error variance will be reduced by its inclusion. Too large residual errors are the main bugbear of forecasting formulae.

As an application has shown, from our data alone we may be able to decide whether the original data or their deltas yield the more efficient estimates of the coefficients.

2. THE RIGHT WAY TO CALCULATE RATE OF CHANGE IN ECONOMIC TIME SERIES

Let the observations be $Z_1, Z_2, \ldots, Z_T$ over a period of "years" (or any other time unit) T. The usual method is to set—

$$(1+r)^{T-1} = Z_T/Z_1, \tag{2.1}$$

where $100r$ is the rate % per year, calculated by logarithms—

$$\log(1+r) = (\log Z_t - \log Z_1)/(T-1), \tag{2.2}$$

r being then found from antilogs.

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

$$Z_T/Z_1 = (Z_2/Z_1)(Z_3/Z_2) \ldots (Z_{T-1}/Z_{T-2})(Z_T/Z_{T-1}), \tag{2.3}$$

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, Z_2 when, having brought it in, one proceeds to cancel it out. In blunt statistical terms this is usually a highly inefficient method of calculation of rate of change.

Assuming that the true rate of change has been constant over all T observations, a more efficient method of calculating it is to apply the classical linear regression model. Assume that the series Z_t, apart from a disturbance term, can be represented by a model $\gamma e^{\beta t}$, β now representing the instantaneous rate of increase, namely $\frac{dZ_t}{Z_t dt}$ at time t, β being significantly different from zero.

In log form, suitable for least squares regression treatment, the model then is—

$$Y_t = \alpha + \beta t + u_t, \quad t = 1, 2, \ldots, T \tag{2.4}$$

where $Y_t = \log_e Z_t$, $\alpha = \log_e \gamma$ and disturbances u_t are assumed to be *regular* (i.e mean zero, homoskedastic-constant variance σ^2 elements mutually uncorrelated). The model is approximately valid (at least enough so for the present purpose) for most Irish economic series in the postwar period. While there may be a tendency for disturbance variance to vary with the magnitude of the data (i.e. the disturbance may be heteroskedastic) this tendency is mitigated by using logs, as in our model.

We shall now compared the relative efficiency of the two methods of estimation of β on stochastic lines, assuming that model (2.4) applies. The first estimate of β, namely b_1, found by taking account only of the first and last terms, is—

$$b_1 = (Y_t - Y_1)/(T-1)$$

$$= \beta + (u_T - u_1)/(T-1), \tag{2.5}$$

from (2.4). Since $E(b_1) = \beta$, b_1 is an unbiased estimate of β. Its variance is—

$$\text{var } b_1 = E(b_1 - \beta)^2 = 2\sigma^2/(T-1)^2 \tag{2.6}$$

The second estimate of β, i.e. the least squares regression estimate b_2 from (2.4) is—

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

$$= \beta + \Sigma u_t(t - \bar{t})/\Sigma(t - \bar{t}).^2 \quad (2.7)$$

$E(b_2) = \beta$, so that b_2 is an unbiased estimate of β. As is also well-known—

$$\text{var } b_2 = E(b_2 - \beta)^2 = \sigma^2/\Sigma(t - \bar{t})^2 \quad (2.8)$$

Now, from the well-known formula for sum squares of first T natural numbers,—

$$\Sigma(t - \bar{t})^2 = T(T^2 - 1)/12. \quad (2.9)$$

Hence—

$$\text{var } b_2 = 12\sigma^2/T(T^2 - 1). \quad (2.10)$$

The relative efficiency of estimate b_1 compared with that of b_2 is defined as—

$$F = \text{var } b_2/\text{var } b_1$$

$$= 6(T - 1)/T(T + 1) \quad (2.11)$$

from (2.6) and (2.10). Hence the estimates are equally efficient (i.e. $F = 1$) for $T = 2$, as they must be; more curiously also when $T = 3$. For large values of T, F is approximately $6/T$.

Although one may have doubts about the validity of model (2.4) as correctly representing the series, the foregoing analysis is so decisive as to lead to the firm recommendation that the first method should never be used for the calculation of rate of change in non-trivial cases. An argument that the wrong method is easier to calculate has no force in this day and age.

Economic and Social Research Institute,
Dublin.