Should Regression in Time Series be Computed Using Original Data or Their Deltas?

All practitioners are aware that least square regression equations in time series are prone to produce r's very near unity because during the regression period (e.g. post-war) all the series tend to have the same time trend, such high correlations being usually regarded as of doubtful significance from the forecasting viewpoint. Two practices are common to avoid the difficulty -

(i) Introduce time t as an indvar;

(ii) Operate the regression on the deltas of the data ($\Delta x_t = x_t - x_{t-1}$)

As to (i), the coefficients of the indvars (except t) are those which would be found if trend t were eliminated from all other variables. In what follows we deal only with (ii). It will suffice to confine attention to the simple (i.e. two-variable) case.

Let the model be

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

where the error term u_t is <u>regular</u> (i.e. for all t $Eu_t = 0$, $Eu^2_t = \sigma^2$, $Eu_t u_{t'}$ $(t' \neq t) = 0$). Then, if b be the regression estimate of β,

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

so that $E\,b = \beta$ with variance

(3) $$\text{var } b = E\,(b - \beta)^2 = \sigma^2 / \Sigma x^2_t,$$

classical results, of course. The Δ version of (1) is

(4) $$Y'_t = \beta X'_t + u'_t, \quad t = 2, 3, \ldots, T,$$

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

Maximum Likelihood

If (1) and (4) (with an additional relation from (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 realization of α and β will be identical. In fact, if the probability element of the vector u is

$$(5) \qquad f(u_1, u_2, \ldots, u_T) \ \Pi\, du_t$$

the ML solution is found as the values of the parameters (e. g. α, β) which maximize f, regarding the u_t in f as functions of the parameters and the data (e. g. as given by (1). If in (5) we make the linear transformation (in matrix form)

$$(6) \qquad u' = A\, u,$$

where A is any non-singular square matrix with numerical elements, (5) transforms into

$$(7) \qquad \left|A^{-1}\right| \ g(u'_1, u'_2, \ldots, u'_T) \ \Pi\, du'_t,$$

where g is the function f after the transformation. Since in (7) the positive determinant $|A|^{-1}$ is a constant (i. e. independent of the parameters), the problem of the maximization of f and g are identical and the maximizing values of the parameters the same. In our particular application the matrix A is

$$(8) A = \begin{vmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{vmatrix}$$

LS Regression Applied to Delta Version (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). Usually a constant term is added, which would indeed be formally consistent with the model:-

(9) $$Y_t = \alpha + \beta X_t + \gamma t + u_t,$$

i.e. as a t (i) in the opening paragraph.

When (4) is regarded as a problem in least squares the estimate b' of β is

(10) $$b' = \beta + \Sigma (x_t - x_{t-1})(u_t - u_{t-1}) / \Sigma (x_t - x_{t-1})^2$$

so that b' is still an unbiased estimate of β.

Its variance,

(11) $$\text{var } b' = 2\sigma^2 \Big(\sum_1^T x'^2_t - \sum_1^T x'_t x'_{t-1} \Big) / \Big(\sum_1^T x'^2_t \Big)^2,$$

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

We cannot compare the efficiency of b and b' for all values of T using the variance formulae (3) and (11) in general algebraic terms so we must have recourse to particular cases.

Case 1

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

(12) $$\Sigma x^2_t = T(T^2 - 1)/12$$

and, from (3)

(13) $$\text{var } b = 12\sigma^2 / T(T^2 - 1).$$

All the x'_t are unity, so that, from (11),

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

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

(15) $$E = 6(T - 1)/T(T + 1)$$

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 to cluster near the median so that our second constructed example will illustrate this. Let there be 2 T observations so that X_t ($t = 1, 2, \ldots, 2T$) 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

(16) $$S_4 = \frac{T}{30}(T + 1)(2T + 1)(3T^2 + 3T - 1),$$

we find, from (3)

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

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 (11) -

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

and efficiency E of estimate b' is

(19) $$E = 5(8T^3 - 2T + 6)^2/6T(T + 1)(2T + 1)(3T^2 + 3T - 1)(4T^2 - 6),$$

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 E is $5/9T'$, 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 (13) and (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 $O(T^0)$. 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 E at (15) and (19) would not be affected.

The General Case

The relative efficiency E is $O(T^{-1})$. If the indvar X_t can be represented by a polynomial of degree r in t, ΣX_t is $O(T^{r+1})$. Using this formula with var b and var b' given by (3) and (11) we find var b $= O(T^{-2T-1})$ and (since $\Delta X_t = x'_t$ is $O(T^{t-1})$)var b' $= O(T^{-2T})$ always $E = O(T^{-1})$ as in Cases 1 and 2.

Consequence of Assumption of u' Regular

Suppose, on the contrary, that in (4) u', by the DW or **r** tests, can be regarded as regular, variance σ^2, what are the implications for Y_t? Clearly -

(20) $$Y_t = \alpha + \beta X_t + u_t$$

but u_t can no longer be regarded as regular, since it is hetero-skedastic. In fact -

(21) $$u_t = \sum_{t'=1}^{t} u'_{t'}$$

so that var $u_t = t\,\sigma^2$. Such a bizarre situation would be outside all experience. Let us, nevertheless follow it through, wrongly assuming that β can be estimated by b'', using LS regression. Then

(22) $$b'' = \Sigma Y_t(X_t - \bar{X}) / \Sigma (X_t - \bar{X})^2$$

which, on substitution from (20), becomes

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

so that $Eb'' = \beta$. However -

(24) $$\hat{\text{var}}\, b'' = \sigma^2 \left(\sum_{t=1}^{T} t x^2_t + 2 \sum_{t=1}^{T} t x_t \sum_{t'=t+1}^{T} x_{t'} \right) / (\Sigma x^2_t)^2$$

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

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 and 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 at 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, T = .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}. The values of the criterion of goodness-of-fit $\Sigma(\Delta Y_t - \Delta Y_{tc})^2$ are 345 and 448, 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 the $\Delta Y_{t'c}$ 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 minimizes this expression.

Conclusion

Estimation of coefficients by LS regression from delta regression is 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 rôle 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 ind-vars 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 act, 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.

18 February 1970 R. C. Geary