THE LOOKOMIC RESEARCH FORTUTEShould Behaviouristic Equation in Time Series be MEMOLAN SERECOmputed in Terms of Absolute Level Variables or their Deltas? NO. 45

be quoted

t o

Not

Confidential:

Authc

the

0 t

permission

the

until

 $\dot{\mathbf{\sigma}}$ Ø

C

obta

Controversy is certain to develop on this issue. In ESRI (Cowling, Leser, O Herlihy) the &'s Practice varies. (i.e. changes in the unit time periods) are generally favoured. Elsewhere (and sometimes in ESRI) the variables themselves are used; and the writer has lately noted a mixed system. left side endogenous being Δ 's and right side being all level variables (e.g. Goldfeld). The writer is unaware of adequate discussion of the problem. Those who favour the \triangle -system are probably influence \hat{a} by the following considerations:-

(i) From the short-term forecasting viewpoint, estimation of change is more important than estimation of absolute level; the absolute value for the previous time is known; hence the change gives the estimate for current time. S (ii) Related to (i): there is acute awareness of the propensity of errors in best-made estimates to be of the Φ tut same order of magnitude as the actual changes.

sti (iii) One wants to use the coefficients of determination $\stackrel{\boldsymbol{\mathsf{L}}}{=} \operatorname{R}^2$ of one's equations to assess their adequacy; the R^2 \mathbf{O} based on absolute level variables are almost useless for this \rightarrow purpose because they are usually so near unity by reason of all $\mathbf{\overline{v}}$ almost constantly going up simultaneously (in the post-war $^{m{\pi}}$ period), an objection which does not necessarily apply to the \triangle variables.

(iv) On a more subtle level: each macro is fairly strongly autoregressed. If one favours simple linear systems (as we do in ESRI, we think advisedly), one knows that the error term in an absolute level type of equation must be a proxy. for many variables (and their non-linearities) which should have been taken into account if they were known. Hence the error term will be prone to be autoregressed. Hence it is not implausible to assume that, u (the model error term) being autoregressed, Au is non-autoregressed. Hence LS regression is more valid as applied to the \triangle form than to the absolute level form of variable.

As a small contribution towards enlightenment, we consider the following problem. In classical simple LS regression, does it matter (from the viewpoint of efficiency of estimation of Δy , as Δy_c) whether we regress on the absolute level variables and then calculate Δy_c from the y_c (Method 1), or whether we set up Δy_c from LS regression of Δy on Δx directly (Method 2). The test of efficiency will be the value of $\Sigma (\Delta y - \Delta y_c)^2$. Let the regression model be

(1) $y_t = \beta x_t + u_t$, $t = 1, 2, \dots, T$, where y_t and x_t are the observations β the unknown population coefficient to be estimated from the data and the constant $\boldsymbol{\epsilon}$ is taken as zero, for simplicity. The error term u_t is a random

(2) $\operatorname{Eu}_{t} = 0$; $\operatorname{Eu}_{t}^{2} = \frac{2}{\sigma}$; $\operatorname{Eu}_{t}u_{t} = 0$, $t' \neq t$.

Method 1

variable, i.e.,

)

The estimate b of β is $\sum \frac{x}{2}$, so that (3) $y_{ct} = x_t \sum \frac{yx}{2x^2} = \beta x_t + \sum \frac{x}{2} \frac{yx}{2} x^2$, using (1). From (1) and (3),

(4) $y_t - y_{ct} = u_t - x_t \sum ux / \sum x^2$ and hence

(5) $y'_t - y'_{ct} = u'_t - x'_t \sum x/\sum x^2$, t = 2, 3, ..., T. Here (for notational simplicity) the primes indicateA 's, i.e. $y'_t = \Delta y_t = y_t - y_{t-1}$ etc. From (2),

(6)
$$\operatorname{Eu}_{t}^{i} = 0; \quad \operatorname{Eu}_{t}^{i}^{2} = 2\sigma^{2}; \quad \operatorname{Eu}_{t}^{i} u_{t+1}^{i} = -\sigma^{2}; \quad \operatorname{Eu}_{t}^{i} u_{t}^{i} = \sigma^{2};$$

 $\operatorname{Eu}_{t}^{i} u_{t+k}^{i} = 0, \quad k > 1.$

Making liberal use of (6), and after some algebra, we find

(7)
$$\frac{1}{2} \sum (y_t^{\circ} - y_{ct}^{\circ})^2 = 2 (T-1) - \Sigma^{\circ} x^{\circ} \frac{2}{2} x^2$$

= 2 (T-1) - d,

where \underline{x} indicates summation from t = 2 to \underline{T} , i.e. $(\underline{T}-1)$ terms and d is the Durbin-Watson (or von Neumann) statistic. (7) gives the efficiency test function for Method 1.

- 2 -

Method 2

Here we start with the model

(8) $y_{t}^{i} = \beta x_{t}^{i} + u_{t}^{i}, t = 2, 3, \dots, T.$

We assume that ρ can be estimated by LS. This assumption is not, of course, correct since the ut are now serially correlated. Actually if the ut are normally distributed the LS solution is the ML solution (and hence asymptotically the most efficient). It is easy to show that the ML solution of (8) (without assuming normality) is identical with that of (1), so that solving (8) properly would yield the same value of the y'ct's and hence of expression (7). In actual practice, we know little about the auto-regressive properties of ut and ut, so we have little compunction about applying LS in both cases, sustained by the fact that, at least, the estimates of ρ under both conditions are consistent.

For simplicity, we use the same notation as in Method 1, though, of course, Method 2 y_{ct}^* is different from that for Method 1. The Method 2 y_{ct}^* is

(9) $y_{ct} = x_t^* \varepsilon \cdot y^* x^* / \varepsilon \cdot x^2$, t=2,3,...,T. On substitution for y' from (8), (9) gives

(10) $y_{ct}^{i} = \rho x^{i} + x_{t}^{i} \varepsilon^{i} u^{i} x^{i} / \varepsilon^{i} x^{i^{2}}$. From (8) and (10),

(11) $y_{t}^{*} - y_{ct}^{*} = u_{t}^{*} - x_{t}^{*} \Sigma' u' x' / \Sigma' x'^{2}$

After much algebra we eventually find

(12) $\frac{1}{c^2} \mathbb{E} (y_t - y_{ct})^2 = 2\mathbb{I} - 4 + 2\Sigma'' x_{t-1}' x_t' / \Sigma' x_t'^2$ where Σ'' indicates summation from t=3 to T. For comparison with (7), which is our ultimate aim, the right side of (12) can be written more suggestively

(13) $\frac{1}{s^2} \ge \sum (y_t^* - y_{ct}^*)^2 = 2 (T-1) - d^* - 2 e/T$ Here d' is the Durbin-Watson statistic for the first differences $x_t^* = \Delta x_t = x_t - x_{t-1}$ while

(14) $2 e/T = (x_2'^2 + x_T'^2)/\Sigma \cdot x_t'^2 0$

The factor 2/T is introduced to indicate the dimension (in T) of

~) ~

the term. Again for purposes of comparison it will be convenient to divide the expressions at (7) and (13) across by 2/(T-1). Finally we have the two expressions

(15)
$$\operatorname{Ez}^{\circ}(y_{t}^{\circ} - y_{ct}^{\circ})^{\frac{2}{2}} \sigma^{2}(T-1) = \begin{cases} \varepsilon_{1}^{2} = 1 - d/2(T-1) \\ \delta_{2}^{2} = 1 - d^{\circ}/2(T-1) - e/T(T-1) \\ \varepsilon_{2}^{2} = 1 - d^{\circ}/2(T-1) - e/T(T-1) \end{cases}$$

according as we use Methods 1 or 2.

According to our way of thinking the <u>whole</u> object of regression is <u>not</u> the estimation of the coefficients (in simple or multiple regression) but the estimation of the dependent variable y_t , i.e. of $y_{c,t}$. (An analogous remark applies to systems of equations). Hence, in a practical application, we regard our exercise as the more successful the smaller the value of the expression on the left side of (15). We are largely unconcerned about probabilistic niceties, as indicated, in particular, by our flagrantly violating them in Method 2. We care little about the well-known optimality of certain methods in the asymptotic case, because we are never in this case in practical applications and our experience with power function analysis had shown that the canons of optimality for indefinitely large samples are not necessarily true for our necessarily small or medium-sized samples (in our case size T).

Comparison of δ_1^2 and δ_2^2 at (15) shows that in general, there is little difference in efficiency between Methods 1 and 2. As we always find, the magnitude depends predominantly on the magnitude of the error variance of σ^2 , to the reduction of which, therefore, our main efforts must be devoted. The difference between δ_1^2 and δ_2^2 arises only in the remaining terms, of order T^{-1} . As to the magnitude of d and d': if the x's or x''s are random to one another their value will be about 2 (independent of T) so that the second term of both expression would be only about 1/(T-1) and they cannot differ much when t is reasonably large; in practice, d and d' are within the range $1 \leq d, d' > 2\frac{1}{2}$. Now,

ì

- 4 -

despite the residual's being autoregressed, $\delta \frac{2}{2}$, derived by LS, must be a minimum and hence $\delta \frac{2}{1}$. What (15) shows is that, in the long run, the difference between $\delta \frac{2}{1}$ and $\delta \frac{2}{2}$ is slight.

Of course this conclusion is based on the simplest case imaginable. We surmise (to put it no higher) that it is also true for single equation multivariate regression and for systems of equations involving 2SLS. As already remarked, the issue does not arise when the full ML method is used, for the solutions are identical.

An Application

2

The data are annual, 1949 to 1965 inclusive. In £million the x_t are gross national expenditure, the y_t money (annual average level), defined in the usual way, i.e. - currency + current deposits. Method 1 regression is

(i) $y_{ct} = 186.21 = 0.2640 (x_t = 643.34), R^2 = .9934, s^2 = (.00552)$ (the bracketed figure is the estimated s.d.) Method 2 regression is

(ii) $y'_{ct} = 10.556 = 0.1891 (x'_t - 41.185)$, $R^2 = .6511$, $s^2 = 24.652$ (0.0370) There is no inconsistency between the coefficients in (i) and (ii): the lower .05 (15 d.f.) limit in (i) is 0.2522 whereas the upper .05 (14 d.f.) limit in (ii) is 0.2685. The ranges overlap though it is a near thing. On account of its lower R^2 the coefficient estimate of 0.1891 at (ii) is not of much value. We show it to 4 places but we are not sure of the population value to 1 place.

As a theory of relationship between money and national expenditure (i) suffers from the fact that the intersept (i.e. the value of y_t when $x_t = 0$) is positive, in fact = 16.34: if there were no expenditure, money could exist. If there were a theory of exact proportionality between the two entities (i.e. no constant term) the LS regression would be

- 5 -

(iii)
$$y_{ct} = 0.2876 x_t$$
, $R^2 = .9850*$

An even simpler theory would be $y_{ct} = x_{t \Sigma} y /_{\Sigma} x$ or

(iv)
$$y_{ct} = 0.2894x_t$$
, $R^2 = .9843*$

High values of R^2 are no guarantees of sufficiently accurate estimates of y_t . For the best estimate (i), the standard error (s) is over 4 which means that, if the formula were used for forecasting money a regular error of this magnitude should be anticipated and **occas**ionally the error might be as high as £6 or £8 million. The formula is worthless for forecasting, since we would do as well with quite naive methods.

The absolute values of $\Sigma^{1}(y_{t} - y_{ct})^{2}$, in which we are primarily interested have the following values Method 1: 448.07, Method 2: 345.13, the latter being the lower because it accrues from LS (formula (ii)). However, the s is seen to be about 5, far too large of the estimation of year to year change in money. Incidentally the von Numanna ratios for formulae (i) and (ii) are respectively 1.75 and 2.23. Neither is significant from the Durbin-Watson table, though superficially they illustrate the improvement in residual antoregression as on moves from absolete levels in time series to their Λ .

12 July 1967

R. C. Geary

* As we are interested only in the accuracy of absolute value estimates then R² have been derived as $1 - \sum (y_t - y_{ct})^2 / \sum (y_t - \overline{y})^2$, for comparison with (i). The actual values (in £ million) of $\sum (y_t - y_{ct})^2$ in the three cases are (i) 256.38, (iii) 590.93, (iv) 618.

- 6 -