THE RESEARCH INSTITUTE MEMORANDUM SERIES NO. 50 (REVISED)

Can Cause and Effect be Distinguished by Simple Regression?

Let data be (X_t, Y_t), $t = 1, 2, \ldots, T$. We assume throughout that T is "large". We try to evolve statistical tests for identifying whether X or Y is the cause, the other variable being the effect. If, in fact, X is the cause, the model is

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

the residue u_t being homoskedastic, with finite variance σ^2 mean zero and randomly ordered, i.e. $E\, u_t u_{t'} = o$, $t' \neq t$. We know that X is the "cause" under these conditions because what (1) says is "knowing the values X_t, α and β I calculate the Y_t by multiplying the X_t by β, adding α and then adding random values u_t to each". The X_t "precede" the Y_t; hence the train of causation. We know that α and β may be estimated by LS regression in an unbiased manner (by a and b) and the estimated residue $u_t = Y_t - a - bX_t$ is a consistent estimate of u_t for each t. Suppose, now, we (erroneously) set up a LS regression of X on Y (i.e. assuming Y is the cause):-

(2) $$X_t = c + dY_t + v_t.$$

The LS regression values of the coefficients c and d will be given by

(3) (i) $c = \bar{X} - d\bar{Y}$

(ii) $d = \Sigma x_t y_t / \Sigma y_t^2$,

where $x_t = X_t - \bar{X}$, $y_t = Y_t - Y = \beta x_t + u_t - \bar{u}$, from (1). Hence by substitution in (3)(ii), with $u_t' = u_t - \bar{u}$,

(4) $$d = \Sigma x_t (\beta x_t + u_t') / \Sigma (\beta x_t + u_t')^2$$
$$= (\beta S^2 + w)/(\beta^2 S^2 + 2\beta w + s^2),$$

where $\Sigma x_t^2 = TS^2$, $\Sigma x_t u_t' = Tw$, $\Sigma u_t^2 = Ts^2$. S^2 and s^2 are ordinary magnitudes while $w = O(T^{-1})$. Hence

(5) $$d \doteq \beta S^2/(\beta^2 S^2 + s^2) = 1/(\beta + s^2/\beta S^2) \doteq 1/(b + s^2/bS^2),$$

The symbol "$\doteq$" meaning "approximately equal to, when T is large. From (5) we infer, at once that

(6) $\quad d < 1/b.$

Clearly this inequality is useless for distinguishing X from Y as causitive variable, i.e. distinguishing b from d, the respective estimates of the coefficients; in fact, (6) is equivalent to $b < 1/d$, reversing the roles of b and d. Nor can we see much hope in the absolute terms a and c: in fact, from (3)(i) and (5).

(7) $\quad c \doteq (s^2\bar{X} / bS^2 - a)/(b + s^2/bS^2),$

which does not seem promising for distinguishing the estimated intercepts a from c.

The last of the more obvious hopes is the von Neumann. If, as already stated, X is the cause then the DW value of

(8) $\quad \hat{u}_t = Y_t - a - bX_t$

will be insignificantly different from 2, indicating residual randomness. Will this be the case if, erroneously, we assume that Y is the independent variable? Then, from (2),

(9)
$$\begin{aligned} v_t &= X_t - c - dY_t \\ &= X_t - c - d\,(\alpha + \beta X_t + u_t) \\ &\doteq X_t - c - d\,(a + bX_t + u_t) \end{aligned}$$

The coefficients of X_t is approximately equal to $(1 - bd)$, calculable from the data. Now, from (5),

(10)
$$\begin{aligned} 1 - bd &\doteq s^2/(b^2S^2 + s^2) \\ &\doteq 1 - R^2, \end{aligned}$$

where R is the coefficient of correlation between X_t and Y_t.

If $R^2 = 1$ exactly then u_t in (1) is zero for all values of t; the relationship between X_t and Y_t is exact; (1) and (2) become absolutley consistent statements, with bd = 1; the X_t term vanishes from the right side of (9);

there is no possibility of distinguishing X from Y as the cause, by this or perhaps any approach. The case of $R^2 = 1$ is however, trivial. It is mentioned merely to indicate the ultimate logic of the von Neumann approach. In general the coefficient of X_t in v_t given by (9) is non-zero. Without loss of generality the X_t can be regarded as arranged in order of magnitude so that the X_t exhibit the phenomenon of serial correlation in marked degree. If follows from (9) that, for a value of $T > r$ the residues v should exhibit significant serial correlation, e.g. a value of DW (v_t) at most equal to the .05 probability critical value. Of course, T, in a particular case may not be large enough to determine significance. If T is large enough the procedure then is as follows. Calculate the two LS regressions. If the residual DW is not significantly different from the first, say of Y on X but significant for X on Y then X is the cause and Y the effect.

The following analysis shows the prospect of distinguishing cause from effect by this DW method as promising. For one thing, one likes to get a large R^2 (since then the calculated relationship is more firm) but clearly the larger the value of R^2 found the larger the value of T required to establish significance. At least there seems to be a good case for calculating both regressions with their residual DWs. Perhaps empirically we may decide on direction of causation from the relative magnitudes of the DWs. With economic time series (e.g. money and income) we are prone to find very large R^2 (sometimes of the order of .99) on the raw data. In these cases the common device of using instead ΔX and ΔY has the effect of reducing R^2 very considerably (one of ten finds R^2 of only .3 or .4). The Δ operation does not change direction of causation, however. So, even if one has only a limited number of pairs of observations, the low value of R^2 found for the Δ data

may not be such an unmixed evil after all as it renders possible identification of the causal variable by DW process. However, we do not pursue this Δ aspect here.

What relation must obtain between S and β (or σ) for the DW procedure proposed to be successful? Let the DW ratio be D. Without loss of generality let X and Y in (1) be regarded as measured from their means so that is exactly, and a and c are approximately zero. Furthermore, let $\beta = 1$, without loss of generality (since the X_t in (1) can be altered proportionately). Then, from (9),

$$(11) \qquad v_t \doteq kx_t - d\hat{u}_t, \quad k \doteq 1 - R^2.$$

Now

$$(12) \qquad D = \Sigma(\Delta v_t)^2 / \Sigma v_t^2$$
$$\doteq k^2 \Sigma (\Delta x_t)^2 + d^2 \Sigma (\Delta u_t)^2 / T(k^2S^2 + d^2s^2).$$

In fact, the numerator of (12) contains a term in $\Sigma \Delta x_t \Delta \hat{u}_t$, which can easily be shown to be $O(T^{\frac{1}{2}})$, hence of lower order than the terms retained, which are $O(T)$. Now, without loss of generality, let the DW ratios of x_t and $\hat{u}_t$ be respectively $e = \Sigma(\Delta x_t)^2/TS^2$ and $m = \Sigma(\Delta\hat{u}_t)^2/Ts^2$

Then

$$(13) \qquad D \doteq (ek^2S^2 + md^2s^2)/(k^2S^2 + d^2s^2).$$

Now $k = 1 - R^2 = s^2/(S^2 + s^2)$ and $d = \Sigma xy/\Sigma y^2 \doteq S^2/(S^2 + s^2)$, recalling that $\beta = 1$. Hence, on substitution of these values in (13) and reducing,

$$(14) \qquad D \doteq \frac{es^2 + mS^2}{s^2 + S^2}$$

If, given T, $D \leqslant D_o$, where D_o is the lower 5% critical value of D on the null-hypothesis, tabled by J. Durbin and G. S. Watson*. The lower limits d_L range from 1.08 for

---

* "Testing for serial correlation in least squares regression. II", Biometrika, vol. 38, parts 1 and 2 (1951).

T = 15 to 1.65 for T = 100. For T = 50, d_L = 1.50, which is convenient. Now for economic time series in the recent period e is very small: for instance, using annual figures for Ireland 1947-1967, the values for log GNP and log money are respectively 0.037 and 0.035. Neglecting therefore the term es^2 in (14) we can identify resi ual serial correlation when

(15) $mS^2 \leqslant 1.5\ (s^2 + S^2)$

or when

(16) $R^2 = \frac{S^2}{(s^2 + S^2)} \leqslant \frac{1.5}{m}$

Since m ≐ 2, the condition is approximately $R^2 \leqslant .75$. Hence if T = 50 we are able to distinguish the causal variable when the correlation coefficient R ⩽ .87, always assuming that one variable deemed causitive or independent has yielded insignificant residual serial correlation.

As the DW table may not always be readily available the following approximate formula for the author's d_L (in our notation D_o) the lower limit of the 5% probability on the null-hypothesis for simple regression.

(17) $D_o = 0.914029 + 0.016350T - 0.00009296T^2$

The approximation is quite good, especially when regard is had to the imprecision of the critical points: thus for T = 15 the values are 1.08 (tabled) 1.14 (17); for T = 50 both are the same at 1.50; for T = 100 the values are 1.65 and 1.62 (17). The formula applies only to the author's range T = 15 to 100. It seems quite likely that DW analysis will also identify the direction also in the case of multivariate LS regression (e.g. we may be interested in identifying causally X and Y as above but have our regression contain additional variables indubitably independent in character); this aspect has not been as yet examined. A particular case of considerable interest is that in which we concern

ourselves with lagged as well as current variables. In fact, this research originated in the question "Is money the cause or the effect of income?" In statistical terms, which of these two models do we prefer:-

$$Y = \alpha + \beta_0 X + \beta_1 X_{-1} + \beta_2 X_{-2} + \dots + u_t \; ; \quad \text{or}$$

$$X = \gamma + \delta_0 Y + \delta_1 Y_{-1} + \delta_2 Y_{-2} + \dots + v_t,$$

X and Y being money and income, or their Δ's? We propose to examine this problem using Irish data for 1947 - 1967 annual and quarterly (seasonally corrected).

June 1968 R. C. Geary

Revised August 1968