Identification of Cause and Effect in Simple Least Squares Regression

R. C. GEARY

GIVEN data (X_t, Y_t, $t=1, 2, \ldots, T$, can we distinguish which variable is causal? If the model be—

$$(1) \qquad Y_t = \alpha + \beta X_t + u_t,$$

with disturbances *regular* (i.e. $Eu_t=0$, $Eu_t^2=\sigma^2$, $Eu_tu_{t'}$, $t' \neq t, =0$, all t, t') then the non-stochastic X is causal (or exogenous) and Y is the effect. The essential character of model (1) from the present point of view is the quasi-independence of X and u; in fact $EX_tu_t = X_tEu_t = 0$, all t. As is usual, our data (X_t, Y_t) are assumed to be a single *realisation* from a possible infinity of samples all with the same X_t, the operation E indicating the arithmetic mean of such infinity.

The "right" regression, namely Y_t on X_t, is—

$$(2) \qquad Y_t = a + bX_t + \hat{u}_t,$$

where LS estimates a and b, and disturbances $\hat{u}_t$ are unbiased and consistent estimates of α, β and u_t respectively. If the model be (1) then the von Neumann statistic for the $\hat{u}_t$ namely—

$$(3) \qquad D_{\hat{u}} = \sum_{t=2}^{T} (\Delta\hat{u}_t)^2 / \sum_{t=1}^{T} \hat{u}_t^2,$$

$\Delta\hat{u}_t = \hat{u}_t - \hat{u}_{t-1}$, will not differ significantly from 2, using the well-known Durbin-Watson approximate probability tables for D_u, indicating that in the population the disturbances are probably non-autoregressed, i.e. that $E\hat{u}_t\hat{u}_{t'}$, $t' \neq t$, $= 0$. In fact when the $\hat{u}_t$ are normal variates $ED\hat{u}=2$, exactly. If theory ordains that X should be the cause of Y and if the von Neumann ratio does not contradict this theory with actual data then the theory might be regarded as proved. It would seem prudent, however, to show also that the theory that the Y_t are the cause of the X_t is untenable. Our method is to examine the von Neumann ratios for the residuals from LS regression both ways, i.e. of Y_t on X_t and of X_t on Y_t. If one is near 2 and the other much less, and if theory does not say us nay, we may confidently accept that we have identified the causal variable.

As the mean of the $\hat{u}_t$, namely $\bar{\hat{u}}_t$, equals zero exactly, from (2),—

$$y_t = bx_t + \hat{u}_t \tag{4}$$

where $y_t = Y_t - \bar{Y}$, $x_t = X_t - \bar{X}$.

The "wrong" regression which we examine is that of X_t on Y_t, namely—

$$X_t = c + dY_t + v_t, \tag{5}$$

Required to estimate residuals v_t in terms of parameters of the right regression when c and d are formally calculated by LS regression, the true relationship being (1).

We have—

$$\text{(i) } c = \bar{X} - d\bar{Y} \tag{6}$$

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

On substitution from (4), (6) (ii) becomes—

$$d = \Sigma x_t(bx_t + \hat{u}_t)/\Sigma(bx_t + \hat{u}_t)^2$$

$$= b\Sigma x_t^2/(b^2\Sigma x_t^2 + \Sigma \hat{u}_t^2) \tag{7}$$

since $\Sigma x_t \hat{u}_t = 0$. Hence, from simple regression theory—

$$bd = \Sigma y^2_{tc}/\Sigma y_t^2$$
$$= r^2, \tag{8}$$

where r is the coefficient of correlation between X_t and Y_t, of course a classical result, From (5)—

$$v_t = X_t - c - dY_t$$

$$= X_t - c - d(a + bX_t + \hat{u}_t)$$

(9) $$= -(c+ad) + (1-bd)X_t - d\hat{u}_t.$$

using (6) (i) and (8). (9) is easily seen to be—

(10) $$v_t = (1-r^2)x_t - r^2\hat{u}_t/b$$

When $r^2=1$, $u_t=0$ exactly for all t, from (8) $bd=1$, from (10) $v_t=0$, all t. This is the only case, trivial of course, in which the two regressions are consistent. (10) also shows that, since $\bar{x}$ and $\bar{\hat{u}}$ $(=\Sigma\hat{u}_t/T)$ are both zero, $\bar{v}$ is also zero, as of course it should be since it is an LS residual. The reader can easily verify that (10) is exactly reversible, i.e.—

(11) $$\hat{u}_t = (1-r^2)y_t - r^2 v_t/d.$$

Another form of (10) is—

(12) $$v_t = x_t - \frac{r^2}{b}y_t.$$

From (10)—

(13) $$\Sigma v_t^2 = (1-r^2)^2\Sigma x_t^2 + r^4\Sigma\hat{u}_t^2/b^2$$

since $\Sigma x_t\hat{u}_t=0$. Also, from (10),—

(14) $$\Sigma(\Delta v_t)^2 = (1-r^2)^2\,\Sigma(\Delta x_t)^2 - 2(1-r^2)r^2\,\Sigma\Delta x_t\Delta\hat{u}_t/b + r^4\,\Sigma(\Delta\hat{u}_t)^2/b^2,$$

where $\Delta v_t = v_t - v_{t-1}$ etc. All Σs in (14) are from $t=2$ to $t=T$. So far, theory has been perfectly general. We assume from now on that T is large. Consider the middle term on the right of (14) and set—

(15) $$z = \frac{1}{T}\sum_{t=2}^{T}\Delta x_t\Delta\hat{u}_t$$

$$= \frac{1}{T}[\Delta x_2(\hat{u}_2 - \hat{u}_1) + \ldots + \Delta x_T(\hat{u}_T - \hat{u}_{T-1})]$$

It is easily seen that $E(z)=0$ and var $z=Ez^2$ is $O(T^{-1})$ in which sense z is $O(T^{-1/2})$.

Other terms divided by T on both sides of (14) are ordinary magnitudes (i.e. $O(T^0)$). Hence the middle term on the right of (14) will be ignored. If the von Neumann ratio for the v_t be D then—

$$D = \Sigma(\Delta v_t)^2/\Sigma v_t^2 \tag{16}$$

If the von Neumann ratios for the x_t and the $\hat{u}_t$ be respectively D_x and D_u and if we set $\Sigma x_t^2 = TS^2$ and $\Sigma \hat{u}_t^2 = Ts^2$ then from (13) and (14)—

$$D \doteq \frac{(1-r^2)^2 S^2 D_x + r^4 s^2 D_u/b^2}{(1-r^2)^2 S + r^4 s^2/b^2} \tag{17}$$

where "$\doteq$" means "approximately equal to", i.e. ignoring terms in $T^{-1/2}$. D can be expressed in simpler form by setting $r^2=b^2S^2/(b^2S^2+s^2)$ and $1-r^2 = s^2/(b^2S^2+s^2)$ in(17) giving—

$$D = \frac{s^2 D_x + b^2 S^2 D_u}{s^2 + b^2 S^2} \tag{18}$$

The von Neumann for the "wrong" regression is given approximately at (18). Will we be able to identify the regression as wrong? If, given T and probability level (·05, ·01, etc.), the right hand side is lower than the Durbin-Watson lower critical value on the null hypothesis namely, in the notation of these authors, d_L, then we can make such an identification. We therefore set—

$$\frac{s^2 D_x + b^2 S^2 D_u}{s^2 + b^2 S^2} \leqslant d_L \tag{19}$$

or

$$\frac{b^2 S^2}{s^2} \leqslant \frac{d_L - D_x}{D_u - d_L} \tag{20}$$

Reverting to r^2 notation, with $r^2 = b^2S^2/(b^2S^2+s^2)$,—

$$r^2 \leqslant \frac{d_L - D_x}{D_u - D_x} \tag{21}$$

This is our basic result. We recall that, because of the approximative character of (17), it also is approximate.

For the von Neumann test, as applied to $\hat{u}_t$ given by (3), to be effective, a particular kind of *ordering* of the original data is implied, as usually happens with time series. In the present case of simple regression (there is no difficulty in dealing analogously with the multi-variate case) this implies that if our first *LS* experiment meant fitting a constant to the data, so that $\hat{u}_t = y_t$, clearly this $\hat{u}_t$ should exhibit the phenomenon of serial correlation for the subsequent test on the "right" $\hat{u}_t$ to show probably absence of autoregression. Otherwise, if, before starting our *LS* regression, we were so unwise as to randomise our original data (i.e. change the "row" sequence, 1, 2, . . . , *T*, to a random sequence) we do not affect any of the familiar *LS* regression values (coefficients and their s.e.'s, *r*, *F*, *s. e. e.*) but we destroy the effectiveness of the von Neumann test with its associated Durbin-Watson null-hypothesis probability theory.

We therefore assume that, our data (here *X* and *Y*) are time series ordered in time, all of which usually exhibit serial correlation. We also assume that, if the model be (1) and we regress Y_t on X_t, in no case will the von Neumann ratio (3) differ significantly from 2.

If the data could be regarded as ordered according to the magnitude of the X_t, D_x in (21) is easily seen to be very small. As an example, if equally spaced sequence of X_t is—

$-n, -(n-1), \ldots, -2, -1, 0, 1, 2, \ldots (n-1), n$, so that $T = 2n+1$ Then—

$$D_x = 2n/[n(n+1)(2n+1)/3]$$

$$= 6/(n+1)(2n+1)$$

$$\doteq 12/T^2,$$

exceedingly small when *T* is large. Or, using actual data for Ireland, in fact annual figures for log GNP and log money 1947–1967, we find values of the von Neumann ratios of 0·037 and 0·035 respectively.

In (21) therefore D_x can be set at zero. Also we take D_u at its average value 2, (21) becomes simply—

$$r^2 \leqslant d_L/2. \tag{22}$$

The Durbin-Watson tables show that as *T* increases d_L increases slowly. Thus for simple regression $d_L = 1{\cdot}50$ for $T = 50$ and $d_L = 1{\cdot}65$ for $T = 100$, so that upper limiting values of r^2, for rejection of hypothesis that *Y* is the cause of *X*, would be respectively ·75 and ·83.

Constructed Example

Mainly to confirm that certain of the approximations we made in the text were valid, and generally to check the algebra, we constructed an example in which the u_t in (1) was a random normal sample with $\sigma^2=1$. The X_t were the sequence—

$$K(-30, -29, \ldots, -2, -1. 0, 1, 2, \ldots, 29, 30),$$

so that $T=61$ and numerical constant K to be determined. Also $\bar{X}=0$, so that $X_t=x_t$, β was taken as 1 and α as 0, i.e. the model was $Y_t=x_t+u_t$, in which the X_t were causal, because the formula shows how the Y_t were derived. In this case the correlation coefficient ρ between the X_t and the Y_t is approximately by $\rho^2=\Sigma x_t^2/(\Sigma x_t^2+\Sigma u_t^2)$, with $\Sigma u_t^2=61$. We found K so that $\Sigma x_t^2=50$ which should yield a value of $\rho=\sqrt{(50/111)}=\cdot67$, certainly significant but not too large, as theory requires. The usual statistics are as follows—

$$T=61.\ \Sigma X_t=0.\ \Sigma X_t^2=50=\Sigma x_t^2$$
$$\Sigma Y_t=-6\cdot52.\ \Sigma Y_t^2=97\cdot4434.\ \Sigma y_t^2=96\cdot7444.$$
$$\Sigma x_t Y_t=46\cdot4368=\Sigma x_t y_t$$
$$b=0\cdot928736.\ a=-0\cdot106885.\ r=\cdot6677$$
$$\Sigma \hat{u}_t^2=53\cdot6190.\ s_u^2=0\cdot9088.$$

By reference to its estimated standard error the estimate b of β (which we know is unity) is on the low side. The value of r is exactly what it should be. The value of $\Sigma(\Delta u_t)^2$ was 112·2730 so that the Durbin-Watson statistic was 2·09, indicating absence of residual auto-regression.

As regards the causally wrong regression of the X_t on the Y_t we find from (13), using the foregoing numerical values,—

$$\Sigma v_t^2 = 27\cdot7131,$$

agreeing to four significant figures with the value calculated directly with the regression. We display the values of the three expression on the right of (14)—

$$\Sigma(\Delta v_t)^2 = \cdot0487-\cdot0085+25\cdot8766$$
$$= 25\cdot9168,$$

the last value agreeing with the value calculated directly from the "wrong" regression. As assumed in the text the value of the middle term is negligible. In deriving relation(14) we also seem justified in neglecting the first term.

The value of the d-statistic is $25\cdot9168/27\cdot7131=0\cdot9352$. This is considerably below the 1 per cent critical value of the 1·38 for $T=60$. The illustration confirms the theory of the text: from our data we have been able to identify the causal variable by rejection of the non-causal.

Economic and Social Research Institute,
Dublin.