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) \(Y_t = a + \beta X_t + u_t\),

developed model regular (i.e. \(E_u = 0, E_u^2 = 0^2, E_{u_t, u_{t'}} = 0, \forall t, t'\)) 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 \(EXu_t = X.Eu_t = 0, \forall t\). As 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) \(Y_t = a + b X_t + u_t\),

where LS estimates \(a\) and \(b\), and disturbances \(u_t\) are unbiased and consistent estimates of \(a\), \(\beta\) and \(u_t\) respectively. If the model be (1) then the von Neumann statistic for the \(u_t\) namely—

(3) \(D_\alpha = \sum_{t=2}^{T} \frac{(\Delta u_t)^2}{\sum_{t=1}^{T} u_t^2}\),

\(\Delta u_t = u_t - u_{t-1}\), will not differ significantly from 2, using the well-known Durbin-Watson approximate probability tables for \(D_\alpha\), indicating that in the
population the disturbances are probably non-autoregressed, i.e. that \( \mathbf{E} \mathbf{u}_t \mathbf{u}_t' = 0 \). In fact when the \( \mathbf{u}_t \) are normal variates \( \mathbf{E} \mathbf{u}_t = 0 \), 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 \( \mathbf{u}_t \), namely \( \mathbf{\overline{u}}_t \), equals zero exactly, from (2),—

\[
y_t = b x_t + \mathbf{u}_t
\]

where \( y_t = Y_t - \mathbf{\overline{Y}} \), \( x_t = X_t - \mathbf{\overline{X}} \).

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

\[
X_t = c + d Y_t + \mathbf{v}_t,
\]

Required to estimate residuals \( \mathbf{v}_t \) 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—

\[
\begin{align*}
(i) & \quad c = \mathbf{\overline{X}} - d \mathbf{\overline{Y}} \\
(ii) & \quad d = \Sigma x_t y_t / \Sigma y_t^2,
\end{align*}
\]

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

\[
d = \Sigma x_t (b x_t + \mathbf{u}_t) / \Sigma (b x_t + \mathbf{u}_t)^2
\]

\[
= b \Sigma x_t^2 / (b^2 \Sigma x_t^2 + \Sigma \mathbf{u}_t^2)
\]

since \( \Sigma x_t \mathbf{u}_t = 0 \). Hence, from simple regression theory—

\[
bd = \Sigma y_t^2 / \Sigma y_t^2
\]

\[
= r^2,
\]

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

\[ v_i = X_i - c - dY_i \]

\[ = X_i - c - d(a + bX_i + \hat{u}_i) \]

(9)

\[ = -(c + ad) + (1 - bd)X_i - d\hat{u}_i. \]

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

(10)

\[ v_i = (1 - r^2)x_i - r^2\hat{u}_i/b \]

When \( r^2 = 1, u_i = 0 \) exactly for all \( t \), from (8) \( bd = 1 \), from (10) \( v_i = 0 \), 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 \( \bar{u} \) \((= \Sigma \hat{u}_i / T)\) are both zero, \( v \) 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)

\[ \hat{u}_i = (1 - r^2)y_i - r^2v_i/d. \]

Another form of (10) is—

(12)

\[ v_i = x_i - r^2y_i. \]

From (10)—

(13)

\[ \Sigma v_i^2 = (1 - r^2)^2 \Sigma x_i^2 + r^4 \Sigma \hat{u}_i^2/b^2 \]

since \( \Sigma x_i \hat{u}_i = 0 \). Also, from (10),—

(14)

\[ \Sigma(\Delta v_i)^2 = (1 - r^2)^2 \Sigma(\Delta x_i)^2 - 2(1 - r^2)r^2 \Sigma \Delta x_i \Delta \hat{u}_i/b + r^4 \Sigma(\Delta \hat{u}_i)^2/b^2 \]

where \( \Delta v_i = v_i - v_{i-1} \) etc. All \( \Sigma \)s 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)

\[ z = \frac{1}{T} \sum_{i=1}^{T} \Delta x_i \Delta \hat{u}_i \]

\[ = \frac{1}{T} \left[ \Delta x_2(\hat{u}_2 - \hat{u}_1) + \ldots + \Delta x_T(\hat{u}_T - \hat{u}_{T-1}) \right] \]
It is easily seen that $E(z) = 0$ and $\text{var } z = Ez^2$ 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 $\nu_i$ be $D$ then—

\begin{equation}
D = \Sigma (A_{\nu_i})^2 / \Sigma \nu_i^2
\end{equation}

If the von Neumann ratios for the $x_i$ and the $u_i$ be respectively $D_x$ and $D_u$ and if we set $\Sigma x_i^2 = TS^2$ and $\Sigma u_i^2 = Ts^2$ then from (13) and (14)—

\begin{equation}
D = \frac{(1-r^2)^2S^2D_x + r^2s^2D_u/b^2}{(1-r^2)^2S + r^2s^2/b^2}
\end{equation}

where "=" 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—

\begin{equation}
D = \frac{s^2D_x + b^2S^2D_u}{s^2 + b^2S^2}
\end{equation}

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 (0.05, 0.01, 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—

\begin{equation}
\frac{s^2D_x + b^2S^2D_u}{s^2 + b^2S^2} \leq d_L
\end{equation}

or

\begin{equation}
\frac{b^2S^2}{s^2} \leq \frac{d_L - D_x}{D_u - D_x}
\end{equation}

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

\begin{equation}
r^2 \leq \frac{d_L - D_x}{D_u - D_x}
\end{equation}
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 \( \hat{u} \) 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 \( \hat{u} = \gamma \), clearly this \( \hat{u} \) should exhibit the phenomenon of serial correlation for the subsequent test on the "right" \( \hat{u} \) 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—

\[
D_x = \frac{2n}{[n(n+1)(2n+1)/3]} \\
= \frac{6}{(n+1)(2n+1)} \\
= 12/T^2,
\]

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 \( D_x \) at its average value 2, (21) becomes simply—

\[
(22) \quad r^2 \leq d_t/2.
\]

The Durbin-Watson tables show that as \( T \) increases \( d_t \) increases slowly. Thus for simple regression \( d_t = 1.50 \) for \( T = 50 \) and \( d_t = 1.65 \) 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 0.75 and 0.83.
**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 \( u \) in (1) was a random normal sample with \( \sigma^2 = 1 \). The \( X_i \) 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 \( X_0 = 0 \), so that \( X_i = x_i, \beta \) was taken as 1 and \( \alpha \) as 0, i.e. the model was \( Y_t = x_t + u_t \), in which the \( X_i \) were causal, because the formula shows how the \( Y_t \) were derived. In this case the correlation coefficient \( \rho \) between the \( X_i \) and the \( Y_t \) is approximately by

\[
\rho^2 = \frac{\Sigma x_i^2}{(\Sigma x_i^2 + \Sigma u_i^2)},
\]

with \( \Sigma u_i^2 = 61 \). We found \( K \) so that \( \Sigma x_i^2 = 50 \) which should yield a value of \( \rho = \sqrt{\frac{50}{111}} \approx .67 \), certainly significant but not too large, as theory requires. The usual statistics are as follows—

\[
T = 61, \quad \Sigma x_i = 0, \quad \Sigma x_i^2 = 50 = \Sigma x_i^4,
\]

\[
\Sigma y_i = -6.52, \quad \Sigma y_i^2 = 97.4434, \quad \Sigma y_i^4 = 96.7444.
\]

\[
\Sigma x_i y_i = 46.4368 = \Sigma x_i y_i^2,
\]

\[
b = 0.928736, \quad a = -0.106885, \quad r = -0.6677
\]

\[
\Sigma u_i^2 = 53.6190, \quad s''_u = 0.9088.
\]

By reference to its estimated standard error the estimate \( b \) of \( \beta \) (which we know is unity) is on the low side. The value of \( r \) is exactly what it should be. The value of \( \Sigma (\Delta u_i)^2 \) 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 \( X_i \) on the \( Y_t \) we find from (13), using the foregoing numerical values,—

\[
\Sigma u_i^2 = 27.7131,
\]

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)—

\[
\Sigma (\Delta u_i)^2 = 0.0487 - 0.0085 + 25.8766
\]

\[
= 25.9168,
\]

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 25.9168/27.7131 = 0.9352. 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._