*Econometrica*, Vol. 37, No. 3 (July, 1969)

# INVESTIGATING CAUSAL RELATIONS BY ECONOMETRIC MODELS AND CROSS-SPECTRAL METHODS

BY C. W. J. GRANGER

There occurs on some occasions a difficulty in deciding the direction of causality between two related variables and also whether or not feedback is occurring. Testable definitions of causality and feedback are proposed and illustrated by use of simple two-variable models. The important problem of apparent instantaneous causality is discussed and it is suggested that the problem often arises due to slowness in recording information or because a sufficiently wide class of possible causal variables has not been used. It can be shown that the cross spectrum between two variables can be decomposed into two parts, each relating to a single causal arm of a feedback situation. Measures of causal lag and causal strength can then be constructed. A generalisation of this result with the partial cross spectrum is suggested.

## 1. INTRODUCTION

THE OBJECT of this paper is to throw light on the relationships between certain classes of econometric models involving feedback and the functions arising in spectral analysis, particularly the cross spectrum and the partial cross spectrum. Causality and feedback are here defined in an explicit and testable fashion. It is shown that in the two-variable case the feedback mechanism can be broken down into two causal relations and that the cross spectrum can be considered as the sum of two cross spectra, each closely connected with one of the causations. The next three sections of the paper briefly introduce those aspects of spectral methods, model building, and causality which are required later. Section 5 presents the results for the two-variable case and Section 6 generalises these results for three variables.

## 2. SPECTRAL METHODS

If $X_t$ is a stationary time series with mean zero, there are two basic spectral representations associated with the series:

(i) the Cramer representation,

$$X_t = \int_{-\pi}^{\pi} e^{it\omega}\, dz_x(\omega), \tag{2.1}$$

where $z_x(\omega)$ is a complex random process with uncorrelated increments so that

$$\begin{aligned} E[dz_x(\omega)\,\overline{dz_x(\lambda)}] &= 0, \qquad \omega \neq \lambda, \\ &= dF_x(\omega), \quad \omega = \lambda; \end{aligned} \tag{2.2}$$

(ii) the spectral representation of the covariance sequence

$$\mu_\tau^{xx} = E[X_t\overline{X}_{t-\tau}] = \int_{-\pi}^{\pi} e^{i\tau\omega}\, dF_x(\omega). \tag{2.3}$$

If $X_t$ has no strictly periodic components, $dF_x(\omega) = f_x(\omega)\,d\omega$ where $f_x(\omega)$ is the power spectrum of $X_t$. The estimation and interpretation of power spectra have been discussed in [**4**] and [**5**]. The basic idea underlying the two spectral representations is that the series can be decomposed as a sum (i.e. integral) of uncorrelated components, each associated with a particular frequency. It follows that the variance of the series is equal to the sum of the variances of the components. The power spectrum records the variances of the components as a function of their frequencies and indicates the relative importance of the components in terms of their contribution to the overall variance.

If $X_t$ and $Y_t$ are a pair of stationary time series, so that $Y_t$ has the spectrum $f_y(\omega)$ and Cramer representation

$$Y_t = \int_{-\pi}^{\pi} e^{it\omega}\,dz_y(\omega),$$

then the cross spectrum (strictly power cross spectrum) $Cr(\omega)$ between $X_t$ and $Y_t$ is a complex function of $\omega$ and arises both from

$$\begin{aligned} E[dz_x(\omega)\,\overline{dz_y(\omega)}] &= 0, & \omega \neq \lambda, \\ &= Cr(\omega)\,d\omega, & \omega = \lambda, \end{aligned}$$

and

$$\mu_\tau^{xy} = E[X_t \overline{Y}_{t-\tau}] = \int_{-\pi}^{\pi} e^{i\tau\omega} Cr(\omega)\,d\omega.$$

It follows that the relationship between two series can be expressed only in terms of the relationships between corresponding frequency components.

Two further functions are defined from the cross spectrum as being more useful for interpreting relationships between variables:

(i) the coherence,

$$C(\omega) = \frac{|Cr(\omega)|^2}{f_x(\omega) f_y(\omega)},$$

which is essentially the square of the correlation coefficient between corresponding frequency components of $X_t$ and $Y_t$, and

(ii) the phase,

$$\phi(\omega) = \tan^{-1} \frac{\text{imaginary part of } Cr(\omega)}{\text{real part of } Cr(\omega)},$$

which measures the phase difference between corresponding frequency components. When one variable is leading the other, $\phi(\omega)/\omega$ measure the extent of the time lag.

Thus, the coherence is used to measure the degree to which two series are related and the phase may be interpreted in terms of time lags.

Estimation and interpretation of the coherence and phase function are discussed in [**4**, Chapters 5 and 6]. It is worth noting that $\phi(\omega)$ has been found to be robust under changes in the stationarity assumption [**4**, Chapter 9].

If $X_t$, $Y_t$, and $Z_t$ are three time series, the problem of possibly misleading correlation and coherence values between two of them due to the influence on both of the third variable can be overcome by the use of partial cross-spectral methods.

The spectral, cross-spectral matrix $\{f_{ij}(\omega)\} = S(\omega)$ between the three variables is given by

$$E\begin{bmatrix} dz_x(\omega) \\ dz_y(\omega) \\ dz_z(\omega) \end{bmatrix} [\overline{dz_x(\omega)}\ \overline{dz_y(\omega)}\ \overline{dz_z(\omega)}] = \{f_{ij}(\omega)\}\, d\omega$$

where

$$\begin{aligned} f_{ij}(\omega) &= f_x(\omega) && \text{when} \quad i = j = x, \\ &= Cr^{xy}(\omega) && \text{when} \quad i = x, j = y, \end{aligned}$$

etc.

The partial spectral, cross-spectral matrix between $X_t$ and $Y_t$ given $Z_t$ is found by partitioning $S(\omega)$ into components:

$$S = \left[\begin{array}{c|c} S_{11} & S_{12} \\ \hline S_{21} & S_{22} \end{array}\right].$$

The partitioning lines are between the second and third rows, and second and third columns. The partial spectral matrix is then

$$S_{xy,z} = S_{11} - S_{12}S_{22}^{-1}S_{21}.$$

Interpretation of the components of this matrix is similar to that involving partial correlation coefficients. Thus, the partial cross spectrum can be used to find the relationship between two series once the effect of a third series has been taken into account. The partial coherence and phase are defined directly from the partial cross spectrum as before. Interpretation of all of these functions and generalisations to the $n$-variable case can be found in [**4**, Chapter 5].

## 3. FEEDBACK MODELS

Consider initially a stationary random vector $X_t = \{X_{1t}, X_{2t}, \ldots, X_{kt}\}$, each component of which has zero mean. A linear model for such a vector consists of a set of linear equations by which all or a subset of the components of $X_t$ are "explained" in terms of present and past values of components of $X_t$. The part not explained by the model may be taken to consist of a white-noise random vector $\varepsilon_t$, such that

$$\begin{aligned} E[\varepsilon_t'\varepsilon_s] &= 0, && t \neq s, \\ &= I, && t = s, \end{aligned} \tag{3.1}$$

where $I$ is a unit matrix and 0 is a zero matrix.

Thus the model may be written as

$$A_0 X_t = \sum_{j=1}^{m} A_j X_{t-j} + \varepsilon_t \tag{3.2}$$

where $m$ may be infinite and the $A$'s are matrices.

The completely general model as defined does not have unique matrices $A_j$ as an orthogonal transformation. $Y_t = \Lambda X_t$ can be performed which leaves the form of the model the same, where $\Lambda$ is the orthogonal matrix, i.e., a square matrix having the property $\Lambda\Lambda' = I$. This is seen to be the case as $\eta_t = \Lambda\varepsilon_t$ is still a white-noise vector. For the model to be determined, sufficient a priori knowledge is required about the values of the coefficients of at least one of the $A$'s, in order for constraints to be set up so that such transformations are not possible. This is the so-called "identification problem" of classical econometrics. In the absence of such a priori constraints, $\Lambda$ can always be chosen so that the $A_0$ is a triangular matrix, although not uniquely, thus giving a spurious causal-chain appearance to the model.

Models for which $A_0$ has nonvanishing terms off the main diagonal will be called "models with instantaneous causality." Models for which $A_0$ has no nonzero term off the main diagonal will be called "simple causal models." These names will be explained later. Simple causal models are uniquely determined if orthogonal transforms such as $\Lambda$ are not possible without changing the basic form of the model. It is possible for a model apparently having instantaneous causality to be transformed using an orthogonal $\Lambda$ to a simple causal model.

These definitions can be illustrated simply in the two variable case. Suppose the variables are $X_t$, $Y_t$. Then the model considered is of the form

$$\begin{aligned} X_t + b_0 Y_t &= \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t', \\ Y_t + c_0 X_t &= \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \varepsilon_t''. \end{aligned} \tag{3.3}$$

If $b_0 = c_0 = 0$, then this will be a simple causal model. Otherwise it will be a model with instantaneous causality.

Whether or not a model involving some group of economic variables can be a simple causal model depends on what one considers to be the speed with which information flows through the economy and also on the sampling period of the data used. It might be true that when quarterly data are used, for example, a simple causal model is not sufficient to explain the relationships between the variables, while for monthly data a simple causal model would be all that is required. Thus, some nonsimple causal models may be constructed not because of the basic properties of the economy being studied but because of the data being used. It has been shown elsewhere [**4**, Chapter 7; **3**] that a simple causal mechanism can appear to be a feedback mechanism if the sampling period for the data is so long that details of causality cannot be picked out.

## 4. CAUSALITY

Cross-spectral methods provide a useful way of describing the relationship between two (or more) variables when one is causing the other(s). In many realistic economic situations, however, one suspects that feedback is occurring. In these situations the coherence and phase diagrams become difficult or impossible to interpret, particularly the phase diagram. The problem is how to devise definitions of causality and feedback which permit tests for their existence. Such a definition was proposed in earlier papers [**4**, Chapter 7; **3**]. In this section, some of these definitions will be discussed and extended. Although later sections of this paper will use this definition of causality they will not completely depend upon it. Previous papers concerned with causality in economic systems [**1**, **6**, **7**, **8**] have been particularly concerned with the problem of determining a causal interpretation of simultaneous equation systems, usually with instantaneous causality. Feedback is not explicitly discussed. This earlier work has concentrated on the form that the parameters of the equations should take in order to discern definite causal relationships. The stochastic elements and the natural time ordering of the variables play relatively minor roles in the theory. In the alternative theory to be discussed here, the stochastic nature of the variables and the direction of the flow of time will be central features. The theory is, in fact, not relevant for nonstochastic variables and will rely entirely on the assumption that the future cannot cause the past. This theory will not, of course, be contradictory to previous work but there appears to be little common ground. Its origins may be found in a suggestion by Wiener [**9**]. The relationship between the definition discussed here and the work of Good [**2**] has yet to be determined.

If $A_t$ is a stationary stochastic process, let $\bar{A}_t$ represent the set of *past* values $\{A_{t-j}, j = 1, 2, \ldots, \infty\}$ and $\bar{\bar{A}}_t$ represent the set of *past and present values* $\{A_{t-j}, j = 0, 1, \ldots, \infty\}$. Further let $\bar{A}(k)$ represent the set $\{A_{t-j}, j = k, k + 1, \ldots, \infty\}$.

Denote the optimum, unbiased, least-squares predictor of $A_t$ using the set of values $B_t$ by $P_t(A|B)$. Thus, for instance, $P_t(X|\bar{X})$ will be the optimum predictor of $X_t$ using only past $X_t$. The predictive error series will be denoted by $\varepsilon_t(A|B) = A_t - P_t(A|B)$. Let $\sigma^2(A|B)$ be the variance of $\varepsilon_t(A|B)$.

The initial definitions of causality, feedback, and so forth, will be very general in nature. Testable forms will be introduced later. Let $U_t$ be all the information in the universe accumulated since time $t - 1$ and let $U_t - Y_t$ denote all this information *apart* from the specified series $Y_t$. We then have the following definitions.

DEFINITION 1: *Causality.* If $\sigma^2(X|U) < \sigma^2(X|\overline{U - Y})$, we say that $Y$ is causing $X$, denoted by $Y_t \Rightarrow X_t$. We say that $Y_t$ is causing $X_t$ if we are better able to predict $X_t$ using all available information than if the information apart from $Y_t$ had been used.

DEFINITION 2: *Feedback.* If

$$\sigma^2(X|\bar{U}) < \sigma^2(X|\overline{U - Y}),$$

$$\sigma^2(Y|\bar{U}) < \sigma^2(Y|\overline{U - X}),$$

we say that feedback is occurring, which is denoted $Y_t \Leftrightarrow X_t$, i.e., feedback is said to occur when $X_t$ is causing $Y_t$ and also $Y_t$ is causing $X_t$.

DEFINITION 3: *Instantaneous Causality*. If $\sigma^2(X|\overline{U}, \overline{\overline{Y}}) < \sigma^2(X|\overline{U})$, we say that instantaneous causality $Y_t \Rightarrow X_t$ is occurring. In other words, the current value of $X_t$ is better "predicted" if the present value of $Y_t$ is included in the "prediction" than if it is not.

DEFINITION 4: *Causality Lag*. If $Y_t \Rightarrow X_t$, we define the (integer) causality lag $m$ to be the least value of $k$ such that $\sigma^2(X|U - Y(k)) < \sigma^2(X|U - Y(k+1))$. Thus, knowing the values $Y_{t-j}, j = 0, 1, \ldots, m-1$, will be of no help in improving the prediction of $X_t$.

The definitions have assumed that only stationary series are involved. In the nonstationary case, $\sigma(X|\overline{U})$ etc. will depend on time $t$ and, in general, the existence of causality may alter over time. The definitions can clearly be generalised to be operative for a specified time $t$. One could then talk of causality existing at this moment of time. Considering nonstationary series, however, takes us further away from testable definitions and this tack will not be discussed further.

The one completely unreal aspect of the above definitions is the use of the series $U_t$, representing *all* available information. The large majority of the information in the universe will be quite irrelevant, i.e., will have no causal consequence. Suppose that all relevant information is numerical in nature and belongs to the vector set of time series $Y_t^D = \{Y_t^i, i \in D\}$ for some integer set $D$. Denote the set $\{i \in D, i \neq j\}$ by $D(j)$ and $\{Y_t^i, i \in D(j)\}$ by $Y_t^{D(j)}$, i.e., the full set of relevant information except one particular series. Similarly, we could leave out more than one series with the obvious notation. The previous definitions can now be used but with $U_t$ replaced by $Y_t$ and $U_t - Y_t$ by $Y^{D(j)}$. Thus, for example, suppose that the vector set consists only of two series, $X_t$ and $Y_t$ and that all other information is irrelevant. Then $\sigma^2(X|\overline{X})$ represents the minimum predictive error variance of $X_t$ using only past $X_t$ and $\sigma^2(X|\overline{X}, \overline{Y})$ represents this minimum variance if both past $X_t$ and past $Y_t$ are used to predict $X_t$. Then $Y_t$ is said to cause $X_t$ if $\sigma^2(X|\overline{X}) > \sigma^2(X|\overline{X}, \overline{Y})$. The definition of causality is now relative to the set $D$. If relevant data has not been included in this set, then spurious causality could arise. For instance, if the set $D$ was taken to consist only of the two series $X_t$ and $Y_t$, but in fact there was a third series $Z_t$ which was causing both within the enlarged set $D' = (X_t, Y_t, Z_t)$, then for the original set $D$, spurious causality between $X_t$ and $Y_t$ may be found. This is similar to spurious correlation and partial correlation between sets of data that arise when some other statistical variable of importance has not been included.

In practice it will not usually be possible to use completely optimum predictors, unless all sets of series are assumed to be normally distributed, since such optimum predictors may be nonlinear in complicated ways. It seems natural to use only linear predictors and the above definitions may again be used under this assumption of linearity. Thus, for instance, the best linear predictor of $X_t$ using only past $X_t$ and past $Y_t$ will be of the form

$$P_t(X|\overline{X}, \overline{Y}) = \sum_{j=1}^{\infty} a_j X_{t-j} + \sum_{j=1}^{\infty} b_j Y_{t-j}$$

where the $a_j$'s and $b_j$'s are chosen to minimise $\sigma^2(X|\overline{X}, \overline{Y})$.

It can be argued that the variance is not the proper criterion to use to measure the closeness of a predictor $P_t$ to the true value $X_t$. Certainly if some other criteria were used it may be possible to reach different conclusions about whether one series is causing another. The variance does seem to be a natural criterion to use in connection with linear predictors as it is mathematically easy to handle and simple to interpret. If one uses this criterion, a better name might be "causality in mean."

The original definition of causality has now been restricted in order to reach a form which can be tested. Whenever the word causality is used in later sections it will be taken to mean "linear causality in mean with respect to a specified set $D$."

It is possible to extend the definitions to the case where a subset of series $D^*$ of $D$ is considered to cause $X_t$. This would be the case if $\sigma^2(X|Y^D) < \sigma^2(X|Y^{D-D^*})$ and then $Y^{D^*} \Rightarrow X_t$. Thus, for instance, one could ask if past $X_t$ is causing present $X_t$. Because new concepts are necessary in the consideration of such problems, they will not be discussed here in any detail.

It has been pointed out already [**3**] that instantaneous causality, in which knowledge of the current value of a series helps in predicting the current value of a second series, can occasionally arise spuriously in certain cases. Suppose $Y_t \Rightarrow X_t$ with lag one unit but that the series are sampled every two time units. Then although there is no real instantaneous causality, the definitions will appear to suggest that such causality is occurring. This is because certain relevant information, the missing readings in the data, have not been used. Due to this effect, one might suggest that in many economic situations an apparent instantaneous causality would disappear if the economic variables were recorded at more frequent time intervals.

The definition of causality used above is based entirely on the predictability of some series, say $X_t$. If some other series $Y_t$ contains information in past terms that helps in the prediction of $X_t$ and if this information is contained in no other series used in the predictor, then $Y_t$ is said to cause $X_t$. The flow of time clearly plays a central role in these definitions. In the author's opinion there is little use in the practice of attempting to discuss causality without introducing time, although philosophers have tried to do so. It also follows from the definitions that a purely deterministic series, that is, a series which can be predicted exactly from its past terms such as a nonstochastic series, cannot be said to have any causal influences other than its own past. This may seem to be contrary to common sense in certain special cases but it is difficult to find a *testable* alternative definition which could include the deterministic situation. Thus, for instance, if $X_t = bt$ and $Y_t = c(t + 1)$, then $X_t$ can be predicted exactly by $b + X_{t-1}$ or by $(b/c)Y_{t-1}$. There seems to be no way of deciding if $Y_t$ is a causal factor of $X_t$ or not. In some cases the notation of the "simplest rule" might be applied. For example, if $X_t$ is some complicated polynomial in $t$ and $Y_t = X_{t+1}$, then it will be easier to predict $X_t$ from $Y_{t-1}$ than from past $X_t$. In some cases this rule cannot be used, as the previous example showed. In any case, experience does not indicate that one should expect economic laws to be simple in nature.

Even for stochastic series, the definitions introduced above may give apparently silly answers. Suppose $X_t = A_{t-1} + \varepsilon_t$, $Y_t = A_t + \eta_t$, and $Z_t = A_t + \gamma_t$, where $\varepsilon_t$,

$\eta_t$, and $\gamma_t$ are all uncorrelated white-noise series with equal variances and $A_t$ is some stationary series. Within the set $D = (X_t, Y_t)$ the definition gives $Y_t \Rightarrow X_t$. Within the set $D' = (X_t, Y_t)$, it gives $Z_t \Rightarrow X_t$. But within the set $D'' = (X_t, Y_t, Z_t)$, neither $Y_t$ nor $Z_t$ causes $X_t$, although the sum of $Y_t$ and $Z_t$ would do so. How is one to decide if either $Y_t$ or $Z_t$ is a causal series for $X_t$? The answer, of course, is that neither is. The causal series is $A_t$ and both $Y_t$ and $Z_t$ contain equal amounts of information about $A_t$. If the set of series within which causality was discussed was expanded to include $A_t$, then the above apparent paradox vanishes. It will often be found that constructed examples which seem to produce results contrary to common sense can be resolved by widening the set of data within which causality is defined.

## 5. TWO-VARIABLE MODELS

In this section, the definitions introduced above will be illustrated using two-variable models and results will be proved concerning the form of the cross spectrum for such models.

Let $X_t$, $Y_t$ be two stationary time series with zero means. The simple causal model is

$$\begin{aligned} X_t &= \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t, \\ Y_t &= \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \eta_t, \end{aligned} \tag{5.1}$$

where $\varepsilon_t, \eta_t$ are taken to be two uncorrelated white-noise series, i.e., $E[\varepsilon_t \varepsilon_s] = 0 = E[\eta_t \eta_s]$, $s \neq t$, and $E[\varepsilon_t \varepsilon_s] = 0$ all $t, s$. In (5.1) $m$ can equal infinity but in practice, of course, due to the finite length of the available data, $m$ will be assumed finite and shorter than the given time series.

The definition of causality given above implies that $Y_t$ is causing $X_t$ provided some $b_j$ is not zero. Similarly $X_t$ is causing $Y_t$ if some $c_j$ is not zero. If both of these events occur, there is said to be a feedback relationship between $X_t$ and $Y_t$. It will be shown later that this new definition of causality is in fact identical to that introduced previously.

The more general model with instantaneous causality is

$$\begin{aligned} X_t + b_0 Y_t &= \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t, \\ Y_t + c_0 X_t &= \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \eta_t. \end{aligned} \tag{5.2}$$

If the variables are such that this kind of representation is needed, then instantaneous causality is occurring and a knowledge of $Y_t$ will improve the "prediction" or goodness of fit of the first equation for $X_t$.

Consider initially the simple causal model (5.1). In terms of the time shift operator $U$—$UX_t = X_{t-1}$—these equations may be written

$$X_t = a(U)X_t + b(U)Y_t + \varepsilon_t, \tag{5.3}$$
$$Y_t = c(U)X_t + d(U)Y_t + \eta_t,$$

where $a(U)$, $b(U)$, $c(U)$, and $d(U)$ are power series in $U$ with the coefficient of $U^0$ zero, i.e., $a(U) = \sum_{j=1}^{m} a_j U^j$, etc.

Using the Cramer representations of the series, i.e.,

$$X_t = \int_{-\pi}^{\pi} e^{it\omega}\, dZ_x(\omega), \qquad Y_t = \int_{-\pi}^{\pi} e^{it\omega}\, dZ_y(\omega),$$

and similarly for $\varepsilon_t$ and $\eta_t$, expressions such as $a(U)X_t$ can be written as

$$a(U)X_t = \int_{-\pi}^{\pi} e^{it\omega} a(e^{-i\omega})\, dZ_x(\omega).$$

Thus, equations (5.3) may be written

$$\int_{-\pi}^{\pi} e^{it\omega}[(1 - a(e^{-i\omega}))\, dZ_x(\omega) - b(e^{-i\omega})\, dZ_y(\omega) - dZ_\varepsilon(\omega)] = 0,$$
$$\int_{-\pi}^{\pi} e^{it\omega}[-c(e^{-i\omega})\, dZ_x(\omega) + (1 - d(e^{-i\omega}))\, dZ_y(\omega) - dZ_\eta(\omega)] = 0,$$

from which it follows that

$$A\begin{bmatrix} dZ_x \\ dZ_y \end{bmatrix} = \begin{bmatrix} dZ_\varepsilon \\ dZ_\eta \end{bmatrix} \tag{5.4}$$

where

$$A = \begin{bmatrix} 1 - a & -b \\ -c & 1 - d \end{bmatrix}$$

and where $a$ is written for $a(e^{-i\omega})$, etc., and $dZ_x$ for $dZ_x(\omega)$, etc.

Thus, provided the inverse of $A$ exists,

$$\begin{bmatrix} dZ_x \\ dZ_y \end{bmatrix} = A^{-1}\begin{bmatrix} dZ_\varepsilon \\ dZ_\eta \end{bmatrix}. \tag{5.5}$$

As the spectral, cross-spectral matrix for $X_t$, $Y_t$ is directly obtainable from

$$E\begin{bmatrix} dZ_x \\ dZ_y \end{bmatrix}[\overline{dZ_x}\, \overline{dZ_y}],$$

these functions can quickly be found from (5.5) using the known properties of $dZ_\varepsilon$ and $dZ_\eta$. One finds that the power spectra are given by

$$f_x(\omega) = \frac{1}{2\pi\Delta}(|1 - d|^2\sigma_\varepsilon^2 + |b|^2\sigma_\eta^2), \tag{5.6}$$
$$f_y(\omega) = \frac{1}{2\pi\Delta}(|c|^2\sigma_\varepsilon^2 + |1 - a|^2\sigma_\eta^2),$$

where $\Delta = |(1 - a)(1 - d) - bc|^2$. Of more interest is the cross spectrum which has the form

$$Cr(\omega) = \frac{1}{2\pi\Delta}[(1 - d)\bar{c}\sigma_\varepsilon^2 + (1 - \bar{a})b\sigma_\eta^2].$$

Thus, the cross spectrum may be written as the sum of two components

$$Cr(\omega) = C_1(\omega) + C_2(\omega), \tag{5.7}$$

where

$$C_1(\omega) = \frac{\sigma_\varepsilon^2}{2\pi\Delta}(1 - d)\bar{c}$$

and

$$C_2(\omega) = \frac{\sigma_\eta^2}{2\pi\Delta}(1 - \bar{a})b.$$

If $Y_t$ is not causing $X_t$, then $b \equiv 0$ and so $C_2(\omega)$ vanishes. Similarly if $X_t$ is not causing $Y_t$ then $c \equiv 0$ and so $C_1(\omega)$ vanishes. It is thus clear that the cross spectrum can be decomposed into the sum of two components—one which depends upon the causality of $X$ by $Y$ and the other on the causality of $Y$ by $X$.

If, for example, $Y$ is not causing $X$ so that $C_2(\omega)$ vanishes, then $Cr(\omega) = C_1(\omega)$ and the resulting coherence and phase diagrams will be interpreted in the usual manner. This suggests that in general $C_1(\omega)$ and $C_2(\omega)$ can each be treated separately as cross spectra connected with the two arms of the feedback mechanism. Thus, coherence and phase diagrams can be defined for $X \Rightarrow Y$ and $Y \Rightarrow X$. For example,

$$C_{\vec{xy}}(\omega) = \frac{|C_1(\omega)|^2}{f_x(\omega)f_y(\omega)}$$

may be considered to be a measure of the strength of the causality $X \Rightarrow Y$ plotted against frequency and is a direct generalisation of coherence. We call $C_{\vec{xy}}(\omega)$ the causality coherence.

Further,

$$\phi_{\vec{xy}}(\omega) = \tan^{-1}\frac{\text{imaginary part of } C_1(\omega)}{\text{real part of } C_1(\omega)}$$

will measure the phase lag against frequency of $X \Rightarrow Y$ and will be called the causality phase diagram.

Similarly such functions can be defined for $Y \Rightarrow X$ using $C_2(\omega)$.

These functions are usually complicated expressions in $a, b, c,$ and $d$; for example,

$$C_{\vec{xy}}(\omega) = \frac{\sigma_\varepsilon^4|(1 - d)c|^2}{(\sigma_\varepsilon^2|1 - d|^2 + \sigma_\eta^2|b|^2)(\sigma_\varepsilon^2|c|^2 + |1 - a|^2\sigma_\eta^2)}.$$

Such formulae merely illustrate how difficult it is to interpret econometric models in terms of frequency decompositions. It should be noted that $0 < |C_{\vec{xy}}(\omega)| < 1$ and similarly for $C_{\vec{yx}}(\omega)$.

As an illustration of these definitions, we consider the simple feedback system

$$\begin{aligned} X_t &= bY_{t-1} + \varepsilon_t, \\ Y_t &= cX_{t-2} + \eta_t, \end{aligned} \tag{5.8}$$

where $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$.

In this case

$$a(\omega) = 0,$$

$$b(\omega) = b\,e^{-i\omega},$$

$$c(\omega) = c\,e^{-2i\omega},$$

$$d(\omega) = 0.$$

The spectra of the series $\{X_t\}$, $\{Y_t\}$ are

$$f_x(\omega) = \frac{1 + b^2}{2\pi|1 - bc\,e^{-3i\omega}|^2}$$

and

$$f_y(\omega) = \frac{1 + c^2}{2\pi|1 - bc\,e^{-3i\omega}|^2},$$

and thus are of similar shape.

The usual coherence and phase diagrams derived from the cross spectrum between these two series are

$$C(\omega) = \frac{c^2 + b^2 + 2bc\cos\omega}{(1 + b^2)(1 + c^2)}$$

and

$$\phi(\omega) = \tan^{-1}\frac{c\sin 2\omega - b\sin\omega}{c\cos 2\omega + b\cos\omega}.$$

These diagrams are clearly of little use in characterising the feedback relationship between the two series.

When the causality-coherence and phase diagrams are considered, however, we get

$$C_{\vec{xy}}(\omega) = \frac{c^2}{(1 + b^2)(1 + c^2)}, \qquad C_{\vec{yx}}(\omega) = \frac{b^2}{(1 + b^2)(1 + c^2)}.$$

Both are constant for all $\omega$, and, if $b \neq 0, c \neq 0$, $\phi_{\vec{xy}}(\omega) = 2\omega$ (time lag of two units),[1] $\phi_{\vec{yx}}(\omega) = \omega$ (time-lag of one unit).

The causality lags are thus seen to be correct and the causality coherences to be reasonable. In particular, if $b = 0$ then $C_{\vec{yx}}(\omega) = 0$, i.e., no causality is found when none is present. (Further, in this new case, $\phi_{\vec{xy}}(\omega) = 0$.)

[1] A discussion of the interpretation of phase diagrams in terms of time lags may be found in Granger and Hatanaka [**4**, Chapter 5].

Other particular cases are also found to give correct results. If, for example, we again consider the same simple model (4.8) but with $\sigma_\varepsilon^2 = 1, \sigma_\eta^2 = 0$, i.e., $\eta_t \equiv 0$ for all $t$, then one finds

$$C_{\overrightarrow{xy}}(\omega) = 1,$$

$$C_{\overrightarrow{yx}}(\omega) = 0,$$

i.e., $X$ is "perfectly" causing $Y$ and $Y$ is not causing $X$, as is in fact the case.

If one now considers the model (5.2) in which instantaneous causality is allowed, it is found that the cross spectrum is given by

$$Cr(\omega) = \frac{1}{2\pi\Delta'}[(1 - d)(\bar{c} - c_0)\sigma_\varepsilon^2 + (1 - \bar{a})(b - b_0)\sigma_\eta^2] \tag{5.9}$$

where

$$\Delta' = |(1 - a)(1 - d) - (b - b_0)(c - c_0)|^2.$$

Thus, once more, the cross spectrum can be considered as the sum of two components, each of which can be associated with a "causality," provided that this includes instantaneous causality. It is, however, probably more sensible to decompose $Cr(\omega)$ into three parts, $Cr(\omega) = C_1(\omega) + C_2(\omega) + C_3(\omega)$, where $C_1(\omega)$ and $C_2(\omega)$ are as in (5.7) but with $\Delta$ replaced by $\Delta'$ and

$$C_3(\omega) = \frac{-1}{2\pi\Delta}[c_0(1 - d)\sigma_\varepsilon^2 + b_0(1 - a)\sigma_\eta^2] \tag{5.10}$$

representing the influence of the instantaneous causality.

Such a decomposition may be useful but it is clear that when instantaneous causality occurs, the measures of causal strength and phase lag will lose their meaning.

It was noted in Section 3 that instantaneous causality models such as (5.2) in general lack uniqueness of their parameters as an orthogonal transformation $\Lambda$ applied to the variables leaves the general form of the model unaltered. It is interesting to note that such transformations do not have any effect on the cross spectrum given by (5.9) or the decomposition. This can be seen by noting that equations (5.2) lead to

$$A\begin{bmatrix} dz_x \\ dz_y \end{bmatrix} = \begin{bmatrix} dz_\varepsilon \\ dz_\eta \end{bmatrix}$$

with appropriate $A$. Applying the transformation $\Lambda$ gives

$$\Lambda A\begin{bmatrix} dz_x \\ dz_y \end{bmatrix} = \Lambda\begin{bmatrix} dz_\varepsilon \\ dz_\eta \end{bmatrix}$$

so that

$$\begin{bmatrix} dz_x \\ dz_y \end{bmatrix} = (\Lambda A)^{-1}\Lambda\begin{bmatrix} dz_\varepsilon \\ dz_\eta \end{bmatrix}$$

$$= A^{-1}\begin{bmatrix} dz_\varepsilon \\ dz_\eta \end{bmatrix}$$

which is the same as if no such transformation had been applied. From its definition, $\Delta$ will possess an inverse. This result suggests that spectral methods are more robust in their interpretation than are simultaneous equation models.

Returning to the simple causal model (5.3),

$$X_t = a(U)X_t + b(U)Y_t + \varepsilon_t,$$
$$Y_t = c(U)X_t + d(U)Y_t + \eta_t,$$

throughout this section it has been stated that $Y_t \not\Rightarrow X_t$ if $b \equiv 0$. On intuitive grounds this seems to fit the definition of no causality introduced in Section 4, within the set $D$ of series consisting only of $X_t$ and $Y_t$. If $b \equiv 0$ then $X_t$ is determined from the first equation and the minimum variance of the predictive error of $X_t$ using past $X_t$ will be $\sigma_\varepsilon^2$. This variance cannot be reduced using past $Y_t$. It is perhaps worthwhile proving this result formally. In the general case, it is clear that $\sigma^2(X|\overline{X}, \overline{Y}) = \sigma_\varepsilon^2$, i.e., the variance of the predictive error of $X_t$, if both past $X_t$ and past $Y_t$ are used, will be $\sigma_\varepsilon^2$ from the top equation. If only past $X_t$ is used to predict $X_t$, it is a well known result that the minimum variance of the predictive error is given by

$$\sigma^2(X|\overline{X}) = \exp \tfrac{1}{2}\pi \int_{-\pi}^{\pi} \log \tfrac{1}{2}\pi\, f_x(\omega)\, d\omega. \tag{5.11}$$

It was shown above in equation (5.6) that

$$f_x(\omega) = \frac{1}{2\pi\Delta}(|1 - d|^2\sigma_\varepsilon^2 + |b|^2\sigma_\eta^2)$$

where $\Delta = |(1 - a)(1 - d) - bc|^2$. To simplify this equation, we note that

$$\int_{-\pi}^{\pi} \log|1 - \alpha\, e^{i\omega}|^2\, d\omega = 0$$

by symmetry. Thus if,

$$f_x(\omega) = \alpha_0 \frac{\pi|1 - \alpha_j\, e^{i\omega}|^2}{\pi|1 - \beta_j\, e^{i\omega}|^2},$$

then $\sigma^2(X|\overline{X}) = \alpha_0$. For there to be no causality, we must have $\alpha_0 = \sigma_\varepsilon^2$. It is clear from the form of $f_x(\omega)$ that in general this could only occur if $|b| \equiv 0$, in which case $2\pi f_x(\omega) = \sigma_\varepsilon^2/|1 - a|^2$ and the required result follows.

## 6. THREE-VARIABLE MODELS

The above results can be generalised to the many variables situation, but the only case which will be considered is that involving three variables.

Consider a simple causal model generalising (5.1):

$$X_t = a_1(U)X_t + b_1(U)Y_t + c_1(U)Z_t + \varepsilon_{1,t},$$
$$Y_t = a_2(U)X_t + b_2(U)Y_t + c_2(U)Z_t + \varepsilon_{2,t},$$
$$Z_t = a_3(U)X_t + b_3(U)Y_t + c_3(U)Z_t + \varepsilon_{3,t},$$

where $a_1(U)$, etc., are polynomials in $U$, the shift operator, with the coefficient of $U^0$ zero. As before, $\varepsilon_{i,t}$, $i = 1, 2, 3$, are uncorrelated, white-noise series and denote the variance $\varepsilon_{i,t} = \sigma_i^2$.

Let $\alpha = a_1 - 1$, $\beta = b_2 - 1$, $\gamma = c_3 - 1$, and

$$A = \begin{bmatrix} \alpha & b_1 & c_1 \\ a_2 & \beta & c_2 \\ a_3 & b_3 & \gamma \end{bmatrix},$$

where $b_1 = b_1(e^{-i\omega})$, etc., as before. Using the same method as before, the spectral, cross-spectral matrix $S(\omega)$ is found to be given by $S(\omega) = A^{-1}k(A')^{-1}$ where

$$k = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}.$$

One finds, for instance, that the power spectrum of $X_t$ is

$$f_x(\omega) = |\Delta|^{-2}[\sigma_1^2|\beta\gamma - c_2 b_3|^2 + \sigma_2^2|c_1 b_3 - \gamma b_1|^2 + \sigma_3^2|b_1 c_2^2 - c_1\beta|^2]$$

where $\Delta$ is the determinant of $A$.

The cross spectrum between $X_t$ and $Y_t$ is

$$C_r^{xy}(\omega) = |\Delta|^{-2}[\sigma_1^2(\beta\gamma - c_2 b_3)\overline{(c_2 a_3 - \gamma a_2)} + \sigma_2^2(c_1 b_3 - b_1\gamma)\overline{(\alpha\gamma - c_1 a_3)} + \sigma_3^2(b_1 c_2 - c_1\beta)\overline{(c_1 a_2 - c_2\alpha)}].$$

Thus, this cross spectrum is the sum of three components, but it is not clear that these can be directly linked with causalities. More useful results arise, however, when partial cross spectra are considered. After some algebraic manipulation it is found that, for instance, the partial cross spectrum between $X_t$ and $Y_t$ given $Z_t$ is

$$C_r^{xy,z}(\omega) = -\frac{[\sigma_1^2\sigma_2^2 b_3 a_3 + \sigma_1^2\sigma_2^2\beta a_2 + \sigma_2^2\sigma_3^2 b_1\alpha]}{f_z'(\omega)}$$

where

$$f_z'(\omega) = \sigma_1^2|\beta\gamma - c_2 b_3|^2 + \sigma_2^2|c_1 b_3 - b_1\gamma|^2 + \sigma_3^2|b_1 c_2 - c_1\beta|^2.$$

Thus, the partial cross spectrum is the sum of three components

$$C_r^{xy,z}(\omega) = C_1^{xy,z} + C_2^{xy,z} + C_3^{xy,z}$$

where

$$C_1^{xy,z} = -\frac{\sigma_1^2\sigma_2^2 b_3 a_3}{f_z'(\omega)}, \quad \text{etc.}$$

These can be linked with causalities. The component $C_1^{xy,z}(\omega)$ represents the interrelationships of $X_t$ and $Y_t$ *through* $Z_t$, and the other two components are direct generalisations of the two causal cross spectra which arose in the two variable case and can be interpreted accordingly.

In a similar manner one finds that the power spectrum of $X_t$, given $Z_t$ is

$$f_{x,z}(\omega) = \frac{\sigma_1^2\sigma_2^2|b_3|^2 + \sigma_1^2\sigma_3^2|\beta|^2 + \sigma_2^2\sigma_3^2|b_1|^2}{f'_z(\omega)}.$$

The causal and feedback relationships between $X_t$ and $Y_t$ can be investigated in terms of the coherence and phase diagrams derived from the second and third components of the partial cross spectrum, i.e.,

$$\text{coherence}\,(\overrightarrow{xy}, z) = \frac{|C_2^{xy,z}|^2}{f_{x,y}f_{y,z}}, \quad \text{etc.}$$

## 7. CONCLUSION

The fact that a feedback mechanism may be considered as the sum of two causal mechanisms and that these causalities can be studied by decomposing cross or partial cross spectra suggests methods whereby such mechanisms can be investigated. Hopefully, the problem of estimating the causal cross spectra will be discussed in a later publication. There are a number of possible approaches and accumulated experience is needed to indicate which is best. Most of these approaches are via the model-building method by which the above results were obtained. It is worth investigating, however, whether a direct method of estimating the components of the cross spectrum can be found.

*The University of Nottingham*

## REFERENCES

[1] BASMAN, R. L.: "The Causal Interpretation of Non-Triangular Systems of Economic Relations," *Econometrica*, Volume 31, 1963, pp. 439–448.

[2] GOOD, I. J.: "A Causal Calculus, I, II," *Brit. J. Philos. Soc.*, Volume 11, 1961, pp. 305–318 and Volume 12, 1962, pp. 43–51.

[3] GRANGER, C. W. J.: "Economic Processes Involving Feedback," *Information and Control*, Volume 6, 1963, pp. 28–48.

[4] GRANGER, C. W. J., AND M. HATANAKA: *Spectral Analysis of Economic Time Series*, Princeton University Press, 1964.

[5] NERLOVE, M.: "Spectral Analysis of Seasonal Adjustment Procedures," *Econometrica*, Volume 32, 1964, pp. 241–286.

[6] ORCUTT, G. H.: "Actions, Consequences and Causal Relations," *Review of Economics and Statistics*, Volume 34, 1952, pp. 305–13.

[7] SIMON, H. A.: "Causal Ordering and Identifiability," *Studies in Econometric Method* (edited by W. C. Hood and T. C. Koopmans), Cowles Commission Monograph 14, New York, 1953.

[8] STROTZ, R. H., AND H. WOLD: "Recursive versus Non-Recursive Systems: An Attempt at Synthesis," *Econometrica*, Volume 28, 1960, pp. 417–427.

[9] WIENER, N.: "The Theory of Prediction," *Modern Mathematics for Engineers*, Series 1 (edited by E. F. Beckenback), Chapter 8, 1956.