A LINEAR THEORY FOR NONCAUSALITY

BY

J. P. FLORENS and M. MOUCHART

TECHNICAL REPORT NO. 51
NOVEMBER 1981

PREPARED UNDER CONTRACT NO0014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH
THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

This document has been approved for public release and sale; its distribution is unlimited.
A LINEAR THEORY FOR NONCAUSALITY

J. P. Florens
CERAMMISE, University of Aix-Marseille
Visiting Stanford University

and

M. Mouchart
CORE, University of Louvain

TECHNICAL REPORT NO. 51

NOVEMBER 1981

PREPARED UNDER CONTRACT NO0014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government.
Approved for public release; distribution unlimited.
A LINEAR THEORY FOR NON-CAUSALITY

J.P. Florens
CERAMMISE, University of Aix-Marseille
Visiting Stanford University

and

M. Mouchart
CORE, University of Louvain

ABSTRACT

Different definitions of non-causality (according to Granger, Sims, Haugh and Pierce,...) are analyzed in terms of orthogonality in the Hilbert space of square integrable variables. Conditions, when necessary, are given for their respective equivalence. Some problems of testability are mentioned. Finally non-causality is also analyzed in terms of "rational expectations", extending previous results of Sims.

Key words: Stochastic processes; non-causality; projections in Hilbert spaces; innovations and rational expectations.
A LINEAR THEORY FOR NON-CAUSALITY

J. P. Florens and M. Mouchart

1.1. Introduction.

Following Granger's (1969) and Sims' (1972) papers, the non-causality concept has taken on great importance in econometrics literature. This concept is essentially the same as the concept of transitivity introduced into statistics by Bahadur (1954) and used in sequential analysis (see e.g. Hall, Wijaman and Gosh (1965)). Intuitively, transitivity can be presented in the following way: a sub-process \((z_n)_n\) of a multivariate stochastic process \((x_n)_n\) is transitive if the past and current values of \(z_n\) are sufficient to forecast \(z_{n+1}\). Equivalently, if \(x_n\) is partitioned into \((z_n, y_n)\), we say that the process generating \(y_n\) does not cause the process generating \(z_n\).

A precise statement of this intuitive definition can be made in different ways. In some of our previous work, non-causality is couched in terms of sequences of independence conditions between \(\sigma\)-fields (see Florens, Mouchart, 1980a, b and Florens, Mouchart, Rolin, 1980). In this paper we propose definitions in terms of sequences of orthogonality conditions between linear subspaces of the Hilbert space of random variables. This kind of presentation is implicit in most of econometrics papers and was explicitly used by Hosoya (1977). (In economics or econometrics literature, see also e.g. Gourieroux, Montfort (1980) or Putia (1981) in which the same kind of mathematical tools are used. In time series literature this kind of presentation is very common. See e.g. Anderson (1971)).
The main purpose of this paper is to show the equivalence of several definitions of non-causality (Granger (1969), Sims (1972, 1980), Haugh and Pierce (1977)). These authors have often simultaneously given a definition of and a test procedure for non-causality. We essentially analyze here the relations between these definitions. Comparison of the properties of test procedures is clearly another story. For example, we shall never use stationarity assumptions in the definitions of non-causality or in the proofs of their equivalence. However, stationarity is a crucial assumption in test procedures.

An important point about non-causality is its relationship with the exogeneity concept used in econometrics literature or, more generally, with the theory of sufficiency and ancillarity in sequential models. This relationship was the main topic of our previous papers (Florens, Mouchart 1980a, Florens, Mouchart, Rolin 1980) in any of which a bibliography can be found. In particular, the relationship with exogeneity is studied in Florens, Mouchart (1980c) and in a paper by Engle, Hendry, Richard (1980). So this point will not be treated here.

This paper is organized in the following way. Notation is presented in the second part of the introduction. Section 2 is devoted to the original definition of Granger and to the main properties of this concept. Sims' first definition, Haugh, Pierce's definition and their respective equivalences to Granger's definition are given in Sections 3 and 4. In Section 5 links between non-causality and rational expectations are pointed out and Sims' second definition is presented and shown to be equivalent to Granger's. Definitions, notation and results on orthogonality in a Hilbert space are recalled in the appendix.
1.2. Notations.

Let \((\Omega, \mathcal{G}, P)\) be a probability space and \(L^2\) be the Hilbert space of square integrable random variables (defined \(P\) – almost surely).

In this paper inner product, orthogonality, projection, completion... are relative to the canonical structure of \(L^2\) (see any book on probability theory e.g. Neveu (1964)). For simplicity we restrict our presentation to a bivariate discrete stochastic process \(x_n = (y_n, z_n)\) \(n=0,1,...\) i.e. to a double sequence of random variables \((y_n)\) and \((z_n)\). All random variables considered are assumed to be elements of \(L^2\). (Note that random variables are defined only almost surely, so we in fact consider a class of stochastic processes such that each is a modification of others).

It must be pointed out that the time index belongs in \(N = \{0,1,...\}\) and not in \(Z = \{-..., -1,0,1,...\}\). (Continuous time is another story!)

This hypothesis does not limit our results and has the advantage of making clear the scale of initial conditions. In fact \(N\) must be completed with a maximum element \(\infty\). If \(Z\) is the time index, it must be completed by a minimum element \(-\infty\) and a maximum element \(\infty\). So \(-\infty\) if the index time is \(Z\), and 0 if the index time is \(N\), play the same role, (in Florens-Nouchart (1980 b) details can be found about this modification of the time index).

Let \((y_n)_{n=0,1,...}\) be a stochastic process. We denote by \(y^n_m (n \leq m)\) the linear subspace of \(L^2\) generated by \(y_n,...,y_m\). (For example, \(y^n_n\) is the subspace generated by \(y_n\)). If \(m\) is finite, such finite-dimensional subspaces are closed. \(y^n_m\) denotes the closed subspace generated by \(y_n, y_{n+1},..., y_m\). \(y^n_0\) represents the history (in the sense of all linear functions of the past) of \(y_{n+1}\). Similar notation is used for \((z_n)\)
2. Granger's Non-Causality

For expository purposes, we first recall Granger's (1969) concept of non-causality along with some of its main properties.

**Definition 2.1.** \( y \) does not linearly cause \( z \) iff

\[
\forall n \geq 0: z_{n+1} \nmid y_0^n | z_0^n + u
\]

In this definition, \( u \) may represent initial conditions and any other relevant information to be used by means of linear functions. Typically \( u \) will include, at least, the constant functions, and also any information available at the start of the process. Information that becomes available later will be introduced in section 5. \( u \) is a closed linear subspace of \( L^2 \).

For instance, when \( u \) contains the constant functions only, \( u \) may be dropped in definition 1 if either of the processes \( y \) or \( z \) have zero-mean. From the definition of conditional orthogonality (see Theorem A.1 in the appendix), condition (2.1) allows several readings. The projection of \( z_{n+1} \) onto the linear space \( y_0^n + z_0^n + u \) is contained in the linear space \( z_0^n + u \). Alternatively, the residual of the projection of \( z_{n+1} \) onto the linear space \( z_0^n + u \) is orthogonal to the linear space \( y_0^n \).

By theorems A.4 and A.8, (2.1) is equivalent to any one of the following properties:

\[
\forall n \geq 0: z_{0}^{n+1} \nmid y_0^n | z_0^n + u
\]
If (2.1) or (2.2) or (2.3) or (2.4) is not satisfied, we shall say that "y linearly causes z". Linear causality and non-causality enjoy several interesting properties.

**Property 2.2.** For any process \( z \), \( z \) does not linearly cause \( z \).

Indeed, by theorem A1,

\[
(2.5) \quad \forall u \, \forall n \geq 0: \, z_{n+1} | z_0^n + u
\]

**Property 2.3.** Linear causality is not transitive.

In other words, \( y \) linearly causes \( z \) and \( z \) linearly causes \( w \) do not together imply \( y \) linearly causes \( w \). (Note that "transitivity" is used here in the usual algebraic sense).

We shall call "\( y \) and \( z \)" the set of processes \((\alpha y_n + \beta z_n)\) for any \( \alpha \) and \( \beta \) in \( \mathbb{R} \). This set may also be viewed as the linear space generated by \( y_n \) and \( z_n \). The history of "\( y \) and \( z \)" up to the instant \( n \) is defined as \( y_0^n + z_0^n \). In other words, "\( y \) and \( z \)" represents the set of information obtainable linearly from the observations \( y_n \) and \( z_n \).
Property 2.4. \( y \) and \( z \) does not linearly cause \( w \) if and only if neither \( y \) nor \( z \) linearly causes \( w \).

In other words, and this is a direct implication of theorem A.3, one has

\[
\forall n \geq 0: \ w_{n+1} \parallel (y^n_0 + z^n_0) \mid w^n_0 + u
\]

if and only if

\[
w_{n+1} \parallel y^n_0 \mid w^n_0 + u \quad \forall n \geq 0
\]

and

\[
w_{n+1} \parallel z^n_0 \mid w^n_0 + u \quad \forall n \geq 0
\]

Property 2.5. \( y \) does not linearly cause \( z \) and \( w \) does not imply that \( y \) does not linearly cause \( z \) (or that \( y \) does not linearly cause \( w \)).

It is therefore possible that \( y \) linearly causes \( z \), \( y \) linearly causes \( w \) but that \( y \) does not linearly cause \( z \) and \( w \). In other words:

\[
\forall n \geq 0, \forall (\alpha, \beta) \in \mathbb{R}^2 : \\
(\alpha w_{n+1} + \beta w_{n+1}) \parallel y^n_0 \mid z^n_0 + w^n_0 + u
\]

does not imply:

\[
z_{n+1} \parallel y^n_0 \mid z^n_0 + u
\]
Property 2.6. \( y \) linearly causes \( z \) and \( w \) does not imply that either \( y \) linearly causes \( z \) or that \( y \) linearly causes \( w \).

It is therefore possible that \( y \) linearly causes \( z \) and \( w \) and that neither \( y \) linearly causes \( z \) nor \( y \) linearly causes \( w \).

Definition 2.1 has suggested testing for non-causality by testing the following property:

\[
(2.8) \quad z_{n+1} \perp y_{n-p} \mid z_{n-q} + u, \quad n > \max(p, q)
\]

for some fixed value of \( p \) and \( q \). In general (2.1) does not imply and is not implied by (2.8). Therefore such a test may be justified only by maintaining some supplementary hypotheses. These may be obtained by means of the following theorem.

Theorem 2.7. Property (2.8) is true for any \( p \) under the following conditions:

\[
(2.1) \quad z_{n+1} \perp y_0 \mid z_0 + u
\]

\[
(2.9) \quad z_{n+1} \perp z_{n-q+1} \mid y_0 + z_{n-q} + u
\]

\[
(2.10) \quad (y_0^n + z_{n-q} + u) \cap (z_0^n + u) = z_{n-q} + u.
\]

Proof. (2.1) and (2.9) imply that:
(2.11) \((y^n_0 + z^n_0 + u) z_{n+1} \in (z^n_0 + u) \cap (y^n_0 + z^n_{n-q} + u)\)

and, by (2.10), the l.h.s. of (2.11) belongs to \((z^n_{n-q} + u)\). i.e.:

(2.12) \((y^n_0 + z^n_0 + u) z_{n+1} \in (z^n_{n-q} + u)\)

i.e.

(2.13) \(z_{n+1} \perp (y^n_0 + z^n_0) \mid z^n_{n-q} + u\).

Clearly (2.13) implies (2.8) for any \(p\).

The role of theorem 2.7 may be viewed as follows. Condition (2.10) means that any linear function of \((y_i, z_j, u: 0 \leq i \leq n, n-q \leq j \leq n)\) that is a.s. equal to a linear function of \((z_i, u: 0 \leq i \leq n)\) is a.s. a function of \((z_i, u: n-q \leq i \leq n)\) only. This condition implies, but is not equivalent, to the following property: the only linear functions on \(y^n_0\) a.s. equal to linear functions on \(z^n_{n-q+1}\) are a.s. equal to linear functions on \((z^n_{n-q} + u)\). Condition (2.10) may be viewed as a linear form of "measurable separability" as defined in Mouchart and Rolin (1979) and may be termed "linear separability". The purpose of this condition is to avoid pathologies that could link the \(y\)-process and the \(z\)-process.

For Gaussian processes, condition (2.9) may be viewed as a Markovian condition of order \(q\) for the conditional process generating \((z_{n+1} \mid y^n_0 + u)\). For general processes, condition (2.9) may be viewed as "linear" Markovian condition on the residual of the projection of \(z_{n+1}\) on \((y^n_0 + u)\).
Theorem 2.7 suggests testing for non-causality under the maintained hypotheses (2.9) and (2.10) by testing (2.8) for some fixed \((p,q)\).

This test is generally performed by testing the significance of the coefficients of \(y_n, \ldots, y_{n-p}\) in the regression of \(z_{n+1}\) on \(y_n, \ldots, y_{n-p}, z_n, \ldots, z_{n-q}, u\) (\(u\) is in this case the constant term). Note that neither (2.9) nor (2.10) involve an assumption of stationarity but stationarity implies that (2.9) may be approximately satisfied for large values of \(q\) (see, e.g., Rozanov (1967)). For autoregressive processes (of order smaller than or equal to \(q\)) condition (2.9) will be exactly satisfied, whether the process is stationary or not.
3. Sims' Non-Causality.

Sims (1972) obtained the following result.

**Theorem 3.1.** $y$ does not linearly cause $z$ if and only if

$$(3.1) \forall n \geq 0 \quad z_0^n \perp y_n | z_0^n + u.$$ 

**Proof.** A general proof of this result was given by Hosoya (1977) but we have found the following proof both simple and insightful.

From (2.2), (2.1) is equivalent to:

$$(3.2) \forall n \geq 0 \quad z_0^{n+1} \perp y_0^n | z_0^n + u.$$ 

This implies (3.1) because, by Theorem A.10, (3.2) is equivalent to:

$$(3.3) \forall n \geq 0 \quad z_0^\infty \perp y_0^n | z_0^n + u.$$ 

Reciprocally by Theorem A3, (3.1) implies:

$$(3.4) \forall n \geq 0, \forall p \geq 0 \quad z_0^\infty \perp y_p^n | z_0^n + u$$

and this implies (3.2) by theorem A.8 and because $y_0^n$ is generated by $(y_p; 0 \leq p \leq n)$. □

The easiest interpretation of 3.1 may be the following: the projection of $y_n$ onto $z_0^\infty + u$ (i.e. the best linear approximation of $y_n$ by an element of $z_0^\infty + u$) belongs in $z_0^n + u$. In other words, only the past and the current values of $z_n$ are relevant to explain $y_n$. 

10
Theorem 3.1 depends crucially on linearity: in terms of independence in probability, this result would be false (See e.g. Florens and Mouchart, 1980b).

An immediate implication of theorem A.11 is the following result.

Theorem 3.2. The following properties are equivalent:

\[(3.5) \quad \forall n \geq 0 \quad z_0 \perp y_n \mid z_0 + y_0^{n-1} + u\]

\[(3.6) \quad \forall n \geq 0 \quad z_0 \perp y_n \mid z_0 + y_0 + u\]

Therefore, (3.5) is an alternative form of linear non-causality if the initial condition \( y_0 \) is a linear function of \( u \). In the non-linear theory, the condition analogous to (3.5) has been introduced as a modified Sims' condition so as to obtain a condition equivalent to Granger's (see e.g., Chamberlain (1980) and Florens and Mouchart (1980a, b and c).

Note that the properties (3.1), (3.5) and (3.6) are not modified if \( z_0 \) is replaced by \( z_{n+1}^\infty \) (theorem A.2).

For practical applications or for hypothesis testing Sims' definition (3.1) can be replaced by:

\[(3.7) \quad \forall n \geq q \quad z_{n+1}^{n+p} \perp y_n \mid z_{n-q}^{n} + u\]

for fixed values of \( p \) and \( q \). In general (3.1) and (3.7) are not
equivalent and further hypotheses are required to guarantee an implication between these two definitions. A theorem analogous to theorem 2.7 could be given without difficulty.

An interesting problem is the relationship between Granger's and Sims' definitions when the dimensions of the future and past of the processes (2.8 and 3.7) are fixed. It should be noted that, in general, there is no relationship between these definitions.

The following theorem shows the equivalence between (2.8) and (3.7) under a condition on the marginal process generating \( (z_n) \).

**Theorem 3.3.** Under the following hypothesis:

\[
\forall n \geq p+q \quad z_{n+1} \perp z_{n-(p+q)} \mid z_{n-q} + u .
\]

The following two conditions are equivalent.

\[
\forall n \geq \max(p,q) \quad z_{n+1} \perp y_{n-p} \mid z_{q} + u
\]

\[
\forall n \geq \max(p,q) \quad z_{n+p+1} \perp y_{n} \mid z_{q} + u .
\]

**Proof.** a) (3.9) \(\Rightarrow\) (3.10) follows from the following property, which will be proved by induction.

\[
\forall n \geq \max(p,q) \quad \forall j = 1, \ldots, p+1 \quad z_{n+j} \perp y_{n} \mid z_{n-q} + u
\]

12
(3.9) implies (3.11) with \( j = 1 \). Let us assume (3.11) is true for any \( j \leq p \). (3.9) gives us:

\[
(3.12) \quad z_{n+j+1} \perp y_{n+j-p}^{n+j} \perp z_{n+j-q}^{n+j} + u.
\]

As \( y_n \perp y_{n+j-p}^{n+j} \), (3.12) implies:

\[
(3.13) \quad z_{n+j+1} \perp y_n \perp z_{n+j-q}^{n+j} + u.
\]

(3.8) implies

\[
(3.14) \quad z_{n+j+1} \perp z_{n-q}^{n+j-q-1} \perp z_{n+j-q}^{n+j} + u
\]

and by theorem A.3, (3.13) and (3.14) imply

\[
(3.15) \quad z_{n+j+1} \perp y_n \perp z_{n-q}^{n+q} + u
\]

or equivalently

\[
(3.16) \quad z_{n+j+1} \perp y_n \perp z_{j+1}^{n+j} + z_{n-q}^{n} + u
\]

and by theorem A.3, (3.11) and (3.16) give:

\[
 z_{n+1}^{n+j+1} \perp y_n \perp z_{n-q}^{n} + u.
\]
b) (3.10) $\implies$ (3.9) follows from the following property, also verified by induction:

\[(3.17) \quad \forall n \geq \max(p,q) \, \forall i = 0, \ldots, p \quad z_{n+1} \downarrow y_{n-i}^{n} \mid z_{n-q}^{n} + u.\]

(3.17) is true for $i = 0$ by (3.10). Let us assume (3.17) is satisfied for any $i \leq p-1$. From (3.10) we get:

\[(3.18) \quad z_{n-i+1}^{n-i+1} \downarrow y_{n-i-1}^{n-i-1} \mid z_{n-i-1-q}^{n-i-1-q} + u.\]

As $z_{n-i}^{n} \subset z_{n-i}^{n-i+p}$, we get from (3.18) (by corollary A.4):

\[(3.19) \quad z_{n-i}^{n-i+p} \downarrow y_{n-i-1}^{n-i-1} \mid z_{n-i-1-q}^{n-i-1-q} + u.\]

\[(3.20) \quad \implies z_{n+1}^{n+1} \downarrow y_{n-i-1}^{n-i-1} \mid z_{n-i-1-q}^{n-i-1-q} + u.\]

By theorem A.3, (3.8) and (3.20) imply:

\[(3.21) \quad z_{n+1}^{n+1} \downarrow y_{n-i-1}^{n-i-1} \mid z_{n-q}^{n} + u.\]

And also by theorem A.3, (3.17) and (3.21) imply:

\[(3.22) \quad z_{n+1}^{n+1} \downarrow y_{n-i-1}^{n} \mid z_{n-q}^{n} + u.\]

(3.17) is then verified and the proof is completed. □
The marginal process generating \((z_n)_{n \geq 0}\) is (linearly) autoregressive if

\[(3.23) \quad \forall n \geq q \quad z_{n+1} \perp z_0^n \mid z_{n-q}^n + u .\]

(3.8) implied by (3.23) but (3.8) is weaker than (3.23). Note that theorem 3.3 cannot be stated in terms of conditional independence instead of conditional orthogonality because the proof depends crucially on (see parts (i)(b) and (ii)(b) of theorem A.3) the fact that, with the same notation as in the appendix, \(E_1 \perp E_2 \mid E_3\) and \(E_1 \perp E_4 \mid E_3\) imply \(E_1 \perp (E_2 + E_4) \mid E_3\). This property has no equivalent in terms of conditional independence (see Florens and Mouchart (1980.b)).

Haugh and Pierce (1977) have suggested analysing the cross-correlations between the innovations of the $z$-process and the innovations of the $y$-process. In this section we compare the approach of Haugh and Pierce and linear non-causality.

In our notation, the innovations of a process $(y_n)_{n \geq 0}$ form the process denoted $y_n - y_0 - y_n = (y_0 - \frac{1}{y_n}) y_n$ i.e. the difference between $y_n$ and its projection on the linear space of all the linear combinations of $y_0, \ldots, y_{n-1}$, or, alternatively, the projection of $y_n$ on the orthogonal complement of this space of linear combinations. So the property stated in the following theorem can be viewed as the Haugh and Pierce definition of non-causality, rewritten in our notation.

**Theorem 4.1.** If $y$ does not linearly cause $z$, then

\[
\forall n \geq 0, \forall p < n \quad (z_{n+1} | z_0^n + u) \perp (y_{p+1} | y_0^p).
\]

**Proof.** Clearly $(y_0^p) \perp y_{p+1} \subset y_0^n$ if $p < n$. Therefore, condition (2.1) implies:

\[
\forall n \geq 0, \forall p < n \quad (z_{n+1} | (y_0^p) \perp y_{p+1} | z_0^n + u).
\]

**Theorem 4.2.** If $y_0 \in z_0^n + u$, then (4.1) implies that $y$ does not linearly cause $z$. 

16
Proof. We first rewrite (4.1) as follows:

(4.3) \[ \forall n \geq 0 \quad (z_0^n + u) \perp z_{n+1} \perp y_{p+1} \mid y_0^p. \]

By theorem A.9, (4.3) is equivalent to:

(4.4) \[ \forall n \geq 0 \quad (z_0^n + u) \perp z_{n+1} \perp y_{p+1} \mid y_0^0. \]

i.e.

(4.5) \[ \forall n \geq 0 \quad (z_{n+1} \mid z_0^n + u) \perp (y_{p+1} \mid y_0^0), \]

and, by theorem A.8, (4.5) is equivalent to

(4.6) \[ \forall n \geq 0 \quad (z_{n+1} \mid z_0^n + u) \perp (y_0^n \mid y_0^0) \]

and (4.6) is equivalent to (2.1), by theorem A.6, if \( y_0 \in z_0^n + u. \)

These theorems show that the equivalence between Haugh and Pierce's condition (4.1) and linear non-causality basically depends on the specification of the initial condition \( y_0 \). If \( u \) has the form \( u = y_0 + v \), then the two approaches are equivalent; otherwise linear non-causality implies, but is not implied, by condition (4.1).
5. **Rational Expectations and Non-Causality.**

Non-causality may be viewed as the condition that the prediction of $z_{n+1}$ based only on its own history $z^n_0$ will not improve if it is also based on the past history of $y$. This suggests that "$y$ does not cause $z$" may be rephrased as "$z$ is self-predictive w.r.t. $y" (for more justification see e.g. Florens and Mouchart (1980c)).

Consider now a sequence of messages and, associated with it, a sequence of "information sets" $I_n$ representing the information contained in all the messages up to instant $n$. Then one may decompose $z_n$ (or $y_n$) into an "expected" component given $z_n$ and an "unexpected" one. An interesting question is to analyze non-causality in terms of such a decomposition for both $y$ and $z$. This was the object of a recent paper by Sims (1980).

In a linear context, the sequence $I_n$ will be an increasing sequence of closed subspaces of $L^2$; often $I_n$ will have the form $w^n_0$ where $\{w_n\}$ is a sequence of "observations". The "expected" component $\hat{z}_n$ of $z_n$ becomes the projection of $z_n$ on $I_n$, and its "unexpected" component $\varepsilon_n$ becomes the projection of $z_n$ on the orthogonal complement of $I_n$. In other words, we have the following decomposition:

\[
(5.1) \quad z_n = \hat{z}_n + \varepsilon_n \quad \hat{z}_n = I_n z_n \quad \varepsilon_n = I_n z_n
\]

Similarly for $y_n$: 

18
Remark. It should be pointed out that a closed subspace is a poor mathematical translation of the intuitive concept of information. Indeed, "knowing" \( w_n \) should involve knowing all (measurable) transformations of \( w_n \), and not only the linear ones. In other words, the \( \sigma \)-field is the natural translation for the concept of information. In the present context, representing expectations by projection on a space of linear functions (defined on \( w_0^0 \)) only may be justified on the grounds of computational simplicity or by a Gaussian assumption. Finally, the natural way to deal with an increasing sequence of \( \sigma \)-fields is to introduce a filtration and to consider as the rational expectation of a given process the nearest process adapted to that filtration (see, e.g. Dellacherie-Meyer, (1976), see also Futia (1981)).

In Sims (1980), the information sets \( I_n \) are taken to be:

\[ I_n = z_0^{n-1} + y_0^{n-1} + u \quad n \geq 1. \]  

Therefore:

\[ e_n = (z_0^{n-1} + y_0^{n-1} + u) \perp z_n \quad n \geq 1 \]
\[ \eta_n = (z_0^{n-1} + y_0^{n-1} + u) \perp y_n \quad n \geq 1. \]

It is also assumed that the initial conditions \( z_0 \) and \( y_0 \) are totally unexpected i.e.
With the above notation, we have the following results.

**Theorem 5.1.** If \( y \) does not linearly cause \( z \), then:

\[
\forall n \geq 0 \quad z_{n+1} \perp \eta_0 \mid \epsilon_0^n + u.
\]

**Proof.** Step 1. We first prove:

\[
\forall n \geq 1 \quad \epsilon_n = (z_0^{n-1} + u) z_n.
\]

We note that condition (2.1) may be written as:

\[
\forall n \geq 1 \quad (y_0^{n-1} + z_0^{n-1} + u) z_n = (z_0^{n-1} + u) z_n.
\]

This is equivalent to (5.8).

Step 2. We now prove:

\[
\epsilon_0^n + u = z_0^n + u \quad n \geq 0.
\]

Given (5.6), this is trivial for \( n = 0 \). Suppose (5.10) is true for some \( n \). Then from (5.8), we have

\[
\epsilon_{n+1} = z_{n+1} - (z_0^n + u) z_{n+1}
\]

which clearly belongs to \( z_0^{n+1} + u \).
Therefore, under (5.10) for some \( n \), \( c_{0}^{n+1} + u \subseteq z_{0}^{n+1} + u \). Conversely, from (5.9) again, \( z_{n+1} = c_{n+1} + (z_{0}^{n} + u) z_{n+1} \) which, under (5.10) for some \( n \), belongs to \( c_{0}^{n+1} + u \); therefore under (5.10) for some \( n \), \( c_{0}^{n+1} + u \supseteq z_{0}^{n+1} + u \).

**Step 3.** We now prove:

\[
(5.11) \quad \eta_{0}^{n} \subseteq y_{0}^{n} + z_{0}^{n} + u \quad \forall n \geq 0
\]

indeed,

\[
\eta_{p} = y_{p} - (y_{0}^{p-1} + z_{0}^{p-1} + u) y_{p} \not\in y_{0}^{n} + z_{0}^{n} + u \quad \forall p \leq n .
\]

The proof is concluded by noticing that (2.1) may be written as:

\[
(5.12) \quad z_{n+1} \perp (y_{0}^{n} + z_{0}^{n} + u) | z_{0}^{n} + u . \quad \square
\]

**Theorem 5.2.** If \( y_{0} \not\in u \), then condition (5.7) implies that \( y \) does not linearly cause \( z \).

**Proof. Step 1.** We first prove that the assumptions imply:

\[
(5.13) \quad \forall n \geq 0 \quad z_{n+1} \perp y_{0}^{n} | c_{0}^{n} + u ,
\]

indeed, (5.7) is equivalent to:
\((e_0^n + u) \downarrow z_{n+1} \downarrow \eta_0^n\) \quad \forall n \geq 0

\Rightarrow (e_0^n + u) \downarrow z_{n+1} \downarrow \eta_p \quad \forall (p, n) \quad 0 \leq p \leq n

\Rightarrow (e_0^n + u) \downarrow z_{n+1} \downarrow (y_0^{p-1} + z_0^{p-1} + u) \downarrow y_p \quad \forall (p, n) \quad 0 \leq p \leq n

\Rightarrow (e_0^n + u) \downarrow z_{n+1} \downarrow y_p \mid y_0^{p-1} + z_0^{p-1} + u \quad \forall (p, n) \quad 0 \leq p \leq n

\Rightarrow (e_0^n + u) \downarrow z_{n+1} \downarrow y_p \mid y_0^0 + z_0^0 + u \quad \text{(theorem A.9)} \quad \forall (p, n) \quad 0 \leq p \leq n

\Rightarrow (e_0^n + u) \downarrow z_{n+1} \downarrow y_p \mid z_0^0 + u \quad \text{(because } y_0^0 \in u) \quad \forall (p, n) \quad 0 \leq p \leq n

\Rightarrow z_{n+1} \downarrow y_p \mid e_0^n + u \quad \forall (p, n) \quad 0 \leq p \leq n

The last step is made by using the fact that \(z_0^0 + u \subseteq e_0^n + u\) and by applying theorem A.6.

**Step 2.** We now prove (2.1) by induction. It is clearly true for \(n = 0\) under the hypothesis \(y_0^0 \in u\). Now suppose (2.1) is true for \(n \leq p\); we prove that (2.1) is also true for \(n = p+1\). From step 1 in theorem 5.1, (5.8) is true for \(n \leq p\) and, from step 2, (5.10) is also true for \(n \leq p\); in particular \(e_0^p + u = z_0^p + u\); therefore, by (5.14), (2.1) is true for \(n = p+1\). \(\square\)
Theorem 5.1 shows that condition (5.7) is, in general, stronger than non-causality but theorems 5.2 shows that condition (5.7) is actually equivalent to non-causality if the initial condition on \( y \) is included in \( u \) (which implies that both expectations \( \hat{z}_n \) and \( \hat{y}_n \) may involve \( y_0 \)).

Extensions of non-causality properties can be easily done in terms of expected or unexpected components of the variables. As an example we give the following version of theorem 3.1.

**Theorem 5.3.** If \( y_0 \not\in u \), \( y \) does not linearly cause \( z \) if and only if

\[
\forall n \geq 0 \quad z_{n+1}^\infty \perp \eta_0^n \mid \epsilon_0^n + u.
\]

**Proof.** Using theorems 5.1 and 5.2 we just have to prove the equivalence of (5.14) and (5.7).

(5.14) implies (5.7) by theorem A.2 (since \( z_{n+1}^\infty \not\in z_{n+1}^\infty \)). (5.7) implies (5.14) by using theorem A.10 (note that if \( y \) does not linearly cause \( z \), we have \( \forall n \quad z_0^n + u = \epsilon_0^n + u \) — see step 2 of the proof of theorem 5.1). \( \square \)

Finally let us note that the property "\( y \) does not cause \( z \)" implies the following conditional orthogonality:

\[
\forall n \geq 0 \quad \forall k \geq 0 \quad z_{n+k+1}^{n+k} \perp \eta_{n+1}^{n+k} \mid \epsilon_{0+k}^{n+k} + \eta_0^n + u
\]

(by theorem 5.1 and corollary A.4). This property is actually the property tested by Sims in his 1980 paper.
Appendix. Orthogonality.

Let $X = (L, \langle \cdot, \cdot \rangle)$ be a Hilbert space on $\mathbb{R}$, i.e. $L$ is a linear space of vectors $x_1$'s and $\langle \cdot, \cdot \rangle$ is an inner product (bilinear, symmetric and positive definite) which makes $L$ complete. Details on this structure can be found in e.g. Halmos (1957) or Greub (1975). The Hilbert space we use in the main body of this paper is the set of (classes of) random variables $y$ defined (up to an almost sure equality) on a probability space $(\Omega, \mathcal{G}, P)$ and such that $E(y^2)$ is finite. The inner product between $y$ and $z$ is then defined by $E(yz)$. However, definitions and results given in this appendix are stated in terms of a general Hilbert space.

Let $E_i$ be complete (or, equivalently, closed) linear subspaces of $L$; where in particular $E_0$ is the subspace containing the null vector only. $E_i + E_j$ denotes the usual sum of subspaces, $x_1 \perp x_2$ denotes the usual orthogonality w.r.t. $\langle \cdot, \cdot \rangle$ (i.e. $\langle x_1, x_2 \rangle = 0$). Likewise $x \perp E$ means $x \perp E \forall e \in E$ and $E_1 \perp E_2$ means $e_1 \perp e_2 \forall e_1 \in E_1$, $e_2 \in E_2$. $E^\perp$ denotes the orthogonal complement of $E$, i.e. $E^\perp = \{ x \in L | x \perp E \}$. Note that $E_0^\perp = L$. Finally, $Ex$ denotes the orthogonal projection of the vector $x$ on the subspace $E$, i.e. the unique vector $e \in E$ such that $(x-e) \perp E$. Note that this is a linear idempotent operation. We shall make use of the following property:

$$\langle x_1, Ex_2 \rangle = \langle Ex_1, x_2 \rangle = \langle Ex_1, Ex_2 \rangle \quad \forall x_1, x_2 \in L \text{ and } \forall E.$$  

Note that $Lx = x$ and $E_0 x = 0$. This notation is extended as follows: $E_2 E_1$ means the projection of $E_1$ on $E_2$. Note that $E_0 E_1 = E_0$ and $LE_1 = E_1$. We shall also use:
We shall introduce in this appendix the two concepts of conditional orthogonality and biconditional orthogonality. They are not new concepts, as they are merely particular cases of orthogonality, but they provide convenient notation and results for our kind of problems.

A.1. Theorem.

The following properties are equivalent and define $E_1$ and $E_2$ are orthogonal conditionally on $E_3$ which is denoted as $E_1 \perp E_2 \mid E_3$:

(i) $\{x_1, x_2\} \in E_1 \times E_2 \Rightarrow (x_1 - E_3 x_1) \perp (x_2 - E_3 x_2)$

(or $E_3 \uparrow E_1 \perp E_3 E_2$ or $E_3 E_1 \perp E_2$ or $E_3 \perp E_2$)

(ii) $x_1 \in E_1 \Rightarrow (E_2 + E_3) x_1 = E_2 x_1$

(or $(E_2 + E_3) E_1 = E_3 E_1$)

(iii) $x_3 \in E_2 \Rightarrow (E_1 + E_3) x_2 = E_3 x_2$

(or $(E_1 + E_3) E_2 = E_3 E_2$).

From (i), conditional orthogonality may be interpreted as the usual orthogonality between the projections of $E_1$ and $E_2$ on $E_3$. In a statistical context, these projections will be recognized as the residuals in regression analysis.
A.2. **Elementary properties.**

(i) \[ E_1 \perp E_2 \mid E_0 = E_1 \perp E_2 \]

(ii) \[ E_4 \subset E_1 \text{ and } E_1 \perp E_2 \mid E_3 \text{ imply } E_4 \perp E_2 \mid E_3 \]

(iii) \[ E_4 \subset E_3 \text{ implies } E_1 \perp E_4 \mid E_3 \forall E_1 \]

(iv) \[ E_4 \subset E_3 \text{ and } E_1 \perp E_2 \mid E_3 \text{ imply } (E_1 + E_4) \perp E_2 \mid E_3 . \]

A.3. **Fundamental property of conditional orthogonality.**

The following properties are equivalent:

(i) (a) \[ E_1 \perp E_2 \mid E_3 \text{ and } (b) E_1 \perp E_4 \mid E_3 \]

(ii) (a) \[ E_1 \perp E_2 \mid E_3 \text{ and } (b) E_1 \perp E_4 \mid E_2 + E_3 \]

(iii) \[ E_1 \perp (E_2 + E_4) \mid E_3 . \]

**Proof.** (iii) = (i) by A.2 iii. To prove that (i) = (iii) take \[ x \in E_2 + E_4, \text{ i.e. } x = x_2 + x_4. \] Then, by the linearity of projection

(i) implies that \( (E_1 + E_3) \cdot x = E_3 \cdot x \). Similarly, (iii) = (ii a) by A.2 iii and (iii) = (ii b) because, by (iii), \( (E_2 + E_3 + E_4) \cdot x_1 = E_3 \cdot x_1 \lor x_1 \in E_1 \) which implies \( (E_2 + E_3) \cdot x_1 = E_3 \cdot x_1 \) as a property of projections. Finally (ii) = (iii) because for any \( x_1 \in E_1 \), \( (E_2 + E_3 + E_4) \cdot x_1 = (E_2 + E_3) \cdot x_1 = E_3 \cdot x_1 \) by (ii b) and (ii a) successively. \( \square \)

\[
E_4 \subseteq E_1 + E_3 \quad \text{and} \quad E_1 \perp E_2 \mid E_3 \quad \text{imply} \quad E_1 \perp E_2 \mid E_4 + E_3 \quad . \quad \Box
\]

A.5. Theorem.

The following properties are equivalent and define "\(E_1\) conditionally on \(E_2\) and \(E_3\) conditionally on \(E_4\) are orthogonal" which is denoted by "\((E_1 \mid E_2) \perp (E_3 \mid E_4)\)"

(i) \( \forall (x_1, x_3) \in E_1 \times E_3 = (x_1 - E_2 x_1) \perp (x_3 - E_4 x_3) \)

(ii) \( E_2 \perp E_1 \perp E_4 \perp E_3 \)

(iii) \( E_1 \perp E_4 \mid E_3 \mid E_2 \)

(iv) \( E_2 \perp E_1 \perp E_3 \mid E_4 \).


(i) \( E_1 \perp E_2 \mid E_3 \Leftrightarrow (E_1 \mid E_3) \perp (E_2 \mid E_3) \)

\[ \text{in particular: } E_1 \perp E_2 \Leftrightarrow (E_1 \mid E_0) \perp (E_2 \mid E_0) \]

(ii) \( E_1 \perp E_2 \mid E_3 \Leftrightarrow (E_1 \mid E_0) \perp (E_2 \mid E_3) \)

\[ \Leftrightarrow E_1 \perp (E_2 \mid E_3) \]

(iii) \( E_3 \subseteq E_1 \quad \text{and} \quad (E_1 \mid E_2) \perp (E_3 \mid E_4) \quad \text{imply} \quad (E_3 \mid E_2) \perp (E_3 \mid E_4) \)

(iv) \( E_4 \subseteq E_2 \quad \text{and} \quad (E_1 \mid E_2) \perp (E_3 \mid E_4) \Leftrightarrow E_1 \perp E_3 \mid E_2 \).
A.7. Fundamental property of biconditional orthogonality.

The following properties are equivalent:

(i)  (a) \((E_1 | E_2) \perp (E_3 | E_4)\) and (b) \((E_1 | E_2) \perp (E_5 | E_4)\)

(ii) (a) \((E_1 | E_2) \perp (E_3 | E_4)\) and (b) \((E_1 | E_2) \perp (E_5 | E_3 + E_4)\)

(iii) \((E_1 | E_2) \perp [(E_3 + E_4) | E_4]\).

Proof. From theorem A.5 (iv) one may replace \((E_1 | E_2)\) by \(\perp_{E_2 E_1}\) and then use A.3. □

A.8. Theorem.

Let \(A\) be a subset of \(L\) and \(E\) be the closed linear subspace of \(L\) generated by \(A\). Then for any closed linear subspaces \(E_1, E_2, E_3\) the following two properties are equivalent:

(i) \(A | E_1 \perp E_2 | E_3\)

(ii) \(E | E_1 \perp E_2 | E_3\).

Sequences of conditional orthogonalities.


Let \((F_n)_{n \geq 0}\) be an increasing sequence \((F_{n-1} \subseteq F_n)\) of closed linear subspaces of \(L\). Then, for any \(E\) and \(G\), the following properties are equivalent:
(i) \( \forall n \geq 0, \quad E \perp F_n \mid F_{n-1} + G \)

(ii) \( \forall n \geq 0, \quad E \perp F_n \mid F_0 + G . \)

**Proof.** (ii) implies (i) by A.4. The converse follows from the property: \( \forall n \geq 0, \forall q = 0, \ldots, n \quad E \perp F_n \mid F_{n-q} + G \) which is proved by induction. It is clearly true for \( q = 1 \). Let us assume this property for general \( q \) and note that (i) implies \( E \perp F_{n-q} \mid F_{n-(q+1)} + G \). The result follows from application of A.3 and A.4. \( \square \)

A.10. **Theorem.**

Let \((E_n)_{n \geq 0}\) and \((F_n)_{n \geq 0}\) be increasing sequences of closed linear subspaces of \( L \) such that \( \forall n \quad F_n \subseteq E_n \). \( E_\infty \) is the closed subspace generated by \( \bigcup_n E_n \). Then the following properties are equivalent:

(i) \( \forall n \geq 0, \quad E_{n+1} \perp F_n \mid E_n + G \)

(ii) \( \forall n \geq 0, \quad E_\infty \perp F_n \mid E_n + G . \)

**Proof.** (ii) implies (i) by A.2. The converse follows by using A.6 and the property:

\( \forall n \geq 0, \forall p \geq 1, \quad E_{n+p} \perp F_n \mid E_n + G . \)

This property is proved by induction. It is true for \( p = 1 \). Let us assume this property is true for general \( p \). (i) implies \( E_{n+p} \perp F_{n+p} \mid E_{n+p} + G \) and A.3 and A.4 implies the result. \( \square \)
A.11. **Theorem.**

With the same definitions as in the preceding theorem the following two properties are equivalent:

(i) \( E_\omega \upharpoonright F_n \upharpoonright E_n + F_{n-1} + C \)

(ii) \( \forall n > 0 \quad E_\omega \upharpoonright F_n \upharpoonright E_n + F_0 + C \).

The proof is essentially the same as the proof of theorem A.6.
References


Office of Naval Research Contract N00014-75-C-0442 (NR-042-034)


A** LINEAR THEORY FOR NONCAUSALITY**

J. P. Florens and M. Mouchart

Department of Statistics, Sequoia Hall
Stanford University
Stanford, CA

Office of Naval Research
Statistics and Probability Program Code 436
Arlington, Virginia 22217

November 1981

32

UNCLASSIFIED

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Stochastic processes; noncausality; projections in Hilbert spaces; innovations and rational expectations.

Different definitions of noncausality (according to Granger, Sims, Haugh and Pierce,...) are analyzed in terms of orthogonality in the Hilbert space of square integrable variables. Conditions, when necessary, are given for their respective equivalence. Some problems of testability are mentioned. Finally noncausality is also analyzed in terms of "rational expectations," extending previous results of Sims.