STATISTICAL REPORT

STATISTICAL CAUSALITY

by

Per Mykland

Report No. 14 July 1986

Department of Mathematics

UNIVERSITY OF BERGEN

Bergen, Norway

Department of Mathematics
University of Bergen
5014 Bergen - U
NORWAY

ISSN - 0333-1865

STATISTICAL CAUSALITY

by

Per Mykland

Report No. 14 July 1986

Abstract

The paper introduces a statistical concept of causality in filtered probability spaces which links nonlinear Granger-causality to setwise causality and to the concept of adapted distribution. The causality concept is shown to be closely connected to extremality of martingale problems. This is applied to the equation $dX_t = a_t(X)dt + dW_t$, and the results are used to derive criteria of when $(X_t^{(2)})$ does not cause (nonlinearly) $X_t^{(1)}$ when the solution (X_t) of the equation is of the form $(X_t^{(1)}, X_t^{(2)})$. These criteria are similar to those which exist for linear causality for time series.

1. Introduction

The purpose of this paper is threefold. We introduce a notion of causality which unifies the nonlinear Granger causality with some related concepts. We then show that this causality is very closely linked to the concept of extremality of measure, and this will then be used to derive testable criteria for noncausality for differential equations.

Linear Granger causality was introduced by Granger (1969), and has since spawned a rich litterature, in particular Sims (1972), Pierce & Haugh (1975), Granger (1980), Tjøstheim (1981), Geweke (1982), Engle et al. (1983) and Florens & Mouchart (1985). We shall study a nonlinear version of the concept. Like the linear one, it defines that the process (Y_t) <u>does not cause</u> the process (X_t) if, for all t, the orthogonal projection of the L^2-space representing X_s, $s > t$, on the space representing X_s and Y_s, $s \leq t$, is contained in the space representing X_s, $s \leq t$. However, the spaces representing stochastic variables are those over the σ-field generated by these variables, while in the linear case they are the smallest closed linear spaces containing the variables. The concept was first suggested in Granger & Newbold (1977), ch. 7.4, and it has since been studied by Chamberlain (1982) and Florens & Mouchart (1982). A similar concept, "local dependence", has been looked into by Schweder (1970) and Aalen et al. (1980). If (X_t, Y_t) is a Markov process, then (X_t) is locally independent of (Y_t) iff (X_t) is Markov and is not caused by (Y_t) in the previously defined (nonlinear) sense (this can easily be seen from Theorem 1 (p.402) in Schweder (1970)).

The study of Granger-causality has been mainly preoccupied with time series. We shall instead concentrate on continuous time processes. Many of the systems to which it is natural to apply tests of causality, take

place in continous time. For example, this is generally the case with an economy. In this case, it may be difficult to capture relations of causality in a discrete-time model, cf. the remarks in Bergstrom (1976), ch. 1.1, 1.2 and 2.1, and the observed "causality" in a discrete-time model may depend on the length of interval between each sampling, see (inter alia) Granger & Hatanaka (1964), ch.7. Using a continous time model may possibly only transfer this problem to the estimation side, but there is reason to hope that it is possible to avoid it altogether by constructing methods for testing causality involving sampling at irregularly (or randomly) spaced times of observation. This might involve something in the line of thinking from Shapiro & Silverman (1960), Beutler (1970), Robinson (1980) and Jones (1981).

The plan of the paper is as follows. Section 2 presents the new concept of causality, and discusses the relationship to nonlinear Granger causality and to studies made in other areas. Section 3 looks at the relationship between causality and martingale problems, and Section 4 applies this to a concrete stochastic differential equation. The results are used in Section 5 to derive testable criteria of noncausality in this equation.

2. Causality and related concepts.

Let (Ω,F,P) be a probablity space. We fix a "time axis" $T(\subseteq \mathbb{R})$ for the space, so that whenever we speak of processes or filtrations, their time parameter describes T. A filtration (F_t) is a family of sub-σ-fields of F which is nondecreasing as a function of t. (G_t) is a subfiltration of (F_t) if $G_t \subseteq F_t$ for all $t \in T$. This will be

denoted $(G_t) \leq (F_t)$. F_∞ is the smallest σ-field containing all the F_ts (even if $\sup T < +\infty$). If (X_t) is a process, F_t^X is the smallest σ-field making X_s, $s \leq t$, measurable. (X_t) is (F_t)-adapted if $(F_t^X) \leq (F_t)$.

We define $\overline{F}^0$ to be the set of $A \in F$ with $P(A) = 0$ or 1. For any σ-field M, $\overline{M} = M \vee \overline{F}^0$ (the smallest σ-field containing both M and $\overline{F}^0$). We say that M is the "smallest (P-a.s.)" σ-field with a certain property if every other σ-field $M' (\subseteq F)$ with the same property satisfies $\overline{M}' \supseteq M$. In the same way, (F_t) is the smallest filtration with a property if any other filtration (G_t) with the same property satisfies $(\overline{G}_t) \geq (F_t)$.

If (F_t), (G_t) and (H_t) are filtrations, we define that (G_t) <u>entirely causes</u> (H_t) <u>within</u> (F_t) <u>relative to</u> P,

$$(H_t) < (G_t) ; (F_t) ; P \qquad (2.1),$$

if (G_t) and (H_t) are subfiltrations of $(\overline{F}_t)$ and

$$\forall t \quad \forall A \in H_\infty \quad P(A|G_t) = P(A|F_t) \qquad (2.2),$$

the references to (F_t) and P being omitted if they are not necessary. (2.2) is the same as saying that H_∞ is conditionally independent of F_t given G_t for all t. For criteria of conditional independence, see Chow & Teicher (1978), and for further theorems, see Mouchart & Rolin (1979). The essence of (2.2) is that all information about (H_t) enters the system (F_t) via (G_t).

We also define that a filtration (H_t) <u>is its own cause</u> (<u>within</u> (F_t)) (<u>relative to</u> P) if $(H_t) < (H_t) ; (F_t) ; P$. The definitions apply to stochastic processes as if we were talking about the corresponding

filtrations. E.g., (X_t) is its own cause within (F_t) if $(F_t^X) < (F_t^X) ; (F_t) ; P$.

A process (X_t) which is its own cause is completely described by its behaviour relative to (F_t^X) . E.g., (X_t) is a Markov process relative to (F_t) iff (X_t) is a Markov process (relative to (F_t^X)) and the process is its own cause. Equivalently a process which is its own cause is characterized by the probability measure on a function space. For example, if C^d is the space of continous R^d-valued functions on $[0,t_0]$, and if $\mathcal{B}(C^d)$ is the Borel-σ-algebra on C^d (under the usual sup-norm), a d-dimensional process (W_t,F_t) is a Wiener-process iff (W_t) is its own cause within (F_t) and (W_t) induces a Wiener-measure on $\mathcal{B}(C^d)$ (see Sect.4).

The connection between nonlinear Granger causality and the concepts defined above, is that (Y_t) does not cause (X_t) iff

$$(F_t^X \vee U) < (F_t^X \vee U) ; (F_t^{X,Y} \vee U) \qquad (2.3) ,$$

U being a σ-field representing initial conditions (in the sense of Florens & Mouchart (1982)).

"Entirely cause" is also connected to the setwise causality studied by Suppes (1970). If (F_t) is a filtration, and M_1 is a sub-σ-field of F_∞ , then at any given time t there is a smallest (P-a.s.) σ-field $M_2 \subseteq F_t$ so that

$$\forall A \in M_1 \quad P(A|F_t) = P(A|M_2) \qquad (2.4) ;$$

the existence of such an M_2 being a corollary to Theorem 4.4 p.II-16 in Mouchart & Rolin (1979). It seems natural to define M_2 as the "cause" of M_1 at time t . If $M_1 = \{A,\tilde{A},\Omega,\Phi\}$, the "cause" M_2 is of a single

event A, and for more complex M_1, M_2 is the smallest (P-a.s.) σ-field containing all the "causes" of the elements of M_1. Although this is not exactly the same model as that studied by Suppes, it is an implementation of the same idea.

This model of causality is closely related to the Bayesian definition of sufficiency. In Bayesian terminology, (2.4) would define M_2 as a sufficient σ-algebra if F_t and M_1 had represented observations and parameters respectively. For further details, see Florens & Mouchart (1982) and Mouchart & Rolin (1979). We emphasize, however, that the above definition of causality does not depend on Bayesianism as it involves no σ-algebras of parameters. The optimal estimates are of future events rather than parameters.

The connection to "entirely cause" is now that $(H_t) < (G_t)$; (F_t) iff $\overline{G}_t$ contains the "cause" of H_∞ at each time t. From the minimality of M_2 it is also seen that there is a smallest (P-a.s.) filtration $(G_t^{(1)})$ which entirely causes (H_t) within (F_t). Also, for $r \geqq 1$, one can define inductively $(G_t^{(r+1)})$ as the smallest (P-a.s.) filtration which entirely causes $(G_t^{(r)})$, and if we set $G_t^{(\infty)}$ to be the smallest σ-field containing all the $G_t^{(r)}$s (for all r), then $(G_t^{(\infty)})$ becomes the smallest (P-a.s.) filtration which is its own cause and which satisfy $(H_t) \leqq (\overline{G}_t^{(\infty)})$.

The $(G_t^{(r)})$s permit us to clarify the connection with the concepts of "adapted distribution" and "synonymity", which in particular have been studied by Aldous (1980), Hoover & Keisler (1984) and Hoover (1984). The two concepts are relations between space/filtration/process-systems of the type $\zeta = (\Omega, F, P, F_t, X_t)$. To define them, let $(G_t^{(r)})$ be defined as above (with $(H_t) = (F_t^X)$ - readers of Hoover & Keisler (1984) will note

that $(\overline{G}_t^{(r)}) = (\overline{F}_t^{m_r X})$ whenever the latter is defined, see Definition 2.16 (p.168) of that paper). Also, let $\hat{\zeta} = (\hat{\Omega},\hat{F},\hat{P},\hat{F}_t,\hat{X}_t)$ be another space/filtration/process-system, and with $(\hat{G}_t^{(r)})$ defined analoguously. It is now easily seen fram Definition 2.6 (p.163) of Hoover & Keisler (1984) that ζ and ζ' have the same adapted distribution iff $(\Omega,\overline{G}_\infty^{(\infty)},P,\overline{G}_t^{(\infty)},X_t)$ and $(\hat{\Omega},\overline{\hat{G}}_\infty^{(\infty)},\hat{P},\overline{\hat{G}}_t^{(\infty)},\hat{X}_t)$ are equal up to null sets; they have the same adapted distribution up to rank r iff $(\Omega,\overline{G}_\infty^{(r)},P,\overline{G}_t^{(r)},X_t)$ and $(\hat{\Omega},\overline{\hat{G}}_\infty^{(r)},\hat{P},\overline{\hat{G}}_t^{(r)},\hat{X}_t)$ are equal up to null sets. Synonymity is to have the same adapted distribution up to rank 1.

To be precise, we also define that two systems $(\Omega,G_\infty,P,G_t,X_t)$ and $(\hat{\Omega},\hat{G}_\infty,\hat{P},\hat{G}_t,\hat{X}_t)$ are equal up to null sets if there is a bijection ϕ between $L^1(G_\infty,P)$ and $L^1(\hat{G}_\infty,\hat{P})$ (in the sense of equivalence classes of stochastic variables) which is also a bijection between $L^1(G_t,P)$ and $L^1(\hat{G}_t,\hat{P})$ for all t, which preserves sums and products, which preserves the expectation $(E = \hat{E}\phi)$, and which satisfies $\phi(X_t) = \hat{X}_t(\hat{P}\text{-a.s.})$ for all t.

Finally, it should be noted that "(H_t) is its own cause" sometimes occurs as a useful assumption in the theory of martingales and stochastic integration, see Brémand & Yor (1978), Yor (1979), and Strook & Yor (1980). We shall see another example of this connection in the next section. In this field, the concept is sometimes known as "Hypothesis (H)". There is also an "Hypothesis (H')"; every (H_t)-martingale is an (F_t)-semimartingale. For this, see in particular Jeulin (1980).

To end this section, we state two results which show some of the basic properties of "entirely cause". They are easily deducted from (2.2), and we have already used statement (i) of the first proposition.

Proposition 2.1.

Let (F_t), (G_t), (H_t) and (I_t) be filtrations on a probability space (Ω,F,P). Assume that

$$(G_t) \leq (F_t) \quad \& \quad (H_t) \leq (F_t) \quad \& \quad (I_t) \leq (F_t) \tag{2.5}$$

Then, all expressions being " ; (F_t); P" ,

i) $$(H_t) < (G_t) \Longrightarrow (H_t) \leq (\overline{G}_t) \tag{2.6}$$

and

ii) $$(H_t) < (G_t) \quad \& \quad (H_t) < (I_t) \Longrightarrow (H_t) \leq (\overline{G}_t \cap \overline{I}_t) \tag{2.7}$$

In particular, (i) together with Proposition 2.2 implies that $<$ is a partial ordering of completed (with $\overline{F}^o$) filtrations.

Proposition 2.2 (alterations in the "framework" filtration).

Let (F_t), (G_t), (H_t) and (I_t) be filtrations on a probability space (Ω,F,P). Then the following statements are equivalent (being "; P").

i) $$(\overline{I}_t) < (\overline{H}_t) \; ; \; (\overline{G}_t) \quad \& \quad (\overline{I}_t) < (\overline{G}_t) \; ; \; (\overline{F}_t) \tag{2.8}$$

ii) $$(\overline{I}_t) < (\overline{H}_t) \; ; \; (\overline{F}_t) \quad \& \quad (\overline{H}_t) \leq (\overline{G}_t) \leq (\overline{F}_t) \tag{2.9}$$

3. Causality and martingale problems.

An extension of a probability space (Ω,F,P) is a probability space $(\hat{\Omega},\hat{F},\hat{P})$ which satisfies that there is a measurable surjective function $f : (\hat{\Omega},\hat{F}) \to (\Omega,F)$ which satisfies $\hat{P}f^{-1} = P$ on F. The

extention must also have the same time axis T as the original space. By convention, if a filtration, say (G_t), is difined on the original space, $(\hat{G}_t)$ is defined as

$$\hat{G}_t = \{f^{-1}(A) : A \in G_t\} \qquad (3.1) .$$

The same applies to processes and sets of processes. There may, however be processes or filtrations bearing the superscript $\hat{}$ without having been defined on the original space. Also, note that $\hat{F} \neq f^{-1}(F)$.

On the space (Ω,F,P), the process (M_t) is an (F_t,P)-martingale if (M_t) is (F_t)-adapted and $M_s = E(M_t|F_s)$ for all $s \geq t$. We then have the following result:

Proposition 3.1.

Let (Ω,G_∞,P) be a probability space with a filtration (G_t). Let H be a set of (G_t,P)-martingales. Then the following statements are equivalent.

i) P is extremal in M, the set of probability measures Q on G_∞ which coincide with P on $G_{-\infty} = \cap_t G_t$, and under which all elements of H are (G_t,Q)- martingales.

ii) For any filtration $(\hat{F}_t)$ on an extension $(\hat{\Omega},\hat{F},\hat{P})$ of (Ω,G_∞,P), if $(\hat{F}_t) \geq (\hat{G}_t)$ and if all the elements of $\hat{H}$ are $(\hat{F}_t,\hat{P})$-martingales, then

$$(\hat{G}_t) < (\hat{G}_t) ; (\hat{F}_t) ; (\hat{P}) \qquad (3.2).$$

Note that the elements of H only have to be defined up to stochastic equivalence under P. This is because $P=a_1P_1+a_2P_2$ $(a_1,a_2>0)$ implies $P_1 << P$.

Proof.

(ii) $\Longrightarrow$ (i) is obvious if $P = a_1P_1 + a_2P_2$ $(a_1, a_2 > 0)$, so that P_1 and P_2 coincide on $G_{-\infty}$ and the elements of H are martingales under P_1, but so that $P_1 \neq P_2$, then set

$$\hat{\Omega} = \Omega \times \{1,2\} \tag{3.3}$$

$$\hat{F}_t = \{A \times \{1\} \cup B \times \{2\} : A, B \in G_t\} \tag{3.4}$$

and

$$\hat{P}(A \times \{1\} \cup B \times \{2\}) = a_1P_1(A) + a_2P_2(B) \tag{3.5}$$

The space $(\hat{\Omega}, \hat{F}_\infty, \hat{P})$ with filtration $(\hat{F}_t)$ satisfies both the conditions of (ii), but contradicts (3.2). Hence the result follows.

(i) $\Longrightarrow$ (ii).

Assume that there is an extension $(\hat{\Omega}, \hat{F}, \hat{P})$ of (Ω, F, P), with a filtration $(\hat{F}_t) \geq (\hat{G}_t)$ so that all the elements of $\hat{H}$ are $(\hat{F}_t, \hat{P})$-martingales, and so that (3.2) is not satisfied. This means that there is a $\hat{K} \in \hat{G}_\infty$, a $u \in T$, and an $\hat{A} \in \hat{F}_u$ for which

$$\int_{\hat{A}} I_{\hat{K}} d\hat{P} \neq \int_{\hat{A}} \hat{E}(I_{\hat{K}} | \hat{G}_u) d\hat{P} \tag{3.6}$$

Define, for $\hat{\eta} \in L^1(\hat{\Omega}, \hat{G}_\infty, \hat{P})$,

$$\phi(\hat{\eta}) = \hat{E}(I_{\hat{A}}\hat{E}(\hat{\eta} | \hat{G}_u) + I_{\tilde{\hat{A}}}\hat{\eta}) \tag{3.7}$$

$\tilde{\hat{A}}$ being the complement of $\hat{A}$. ϕ is obviously linear, and by Jensen's inequality it is bounded:

$$|\phi(\hat{\eta})| \leq 2E|\hat{\eta}| \tag{3.8}$$

As $\phi(\hat{\eta}) \geq 0$ for $\hat{\eta} \geq 0$, and as $\phi(1) = 1$, ϕ is the expectation operator

of a probability measure $\hat{Q}$ on $\hat{G}_\infty$:

$$\forall\ \hat{C} \in \hat{G}_\infty \quad \hat{Q}(\hat{C}) = \phi(I_{\hat{C}}) \tag{3.9}$$

$\hat{Q} \neq \hat{P}$ on $\hat{G}_\infty$ since $\hat{Q}(\hat{K}) \neq \hat{P}(\hat{K})$.

Define $\hat{R}$ by $\hat{P} = \frac{1}{2}\hat{Q} + \frac{1}{2}\hat{R}$. By (3.8), $\hat{R}$ is a probability measure on $\hat{G}_\infty$, and $\hat{Q}$ and $\hat{R}$ obviously coincide with $\hat{P}$ on $\hat{G}_{-\infty}$. We shall show that all elements of $\hat{H}$ are $(\hat{F}_t,\hat{Q})$-martingales. Hence they are also $(\hat{F}_t,\hat{R})$-martingales. It follows (by an ordinary change-of-variables formula, see, e.g., Royden (1968)s Theorem 15.1) that $\hat{Q}f^{-1}$ and $\hat{R}f^{-1} \in M$, hence P is not extremal in M . This reduces the assumption that (ii) is not satisfied to absurdity, and the result is proved.

Let $(\hat{N}_t) \in \hat{H}$, let $s \in T$, let $\hat{B} \in \hat{G}_s$. Let $s \leq u$. Since $(\hat{N}_t)$ is an $(\hat{F}_t,\hat{P})$-martingale,

$$\hat{E}(\hat{N}_\infty | \hat{F}_u) = \hat{E}(\hat{N}_\infty | \hat{G}_u) \tag{3.10}$$

whence, as $\hat{A} \in \hat{F}_u$ and $B \in \hat{G}_u \subset \hat{F}_u$,

$$I_{\hat{A}}\hat{E}(I_{\hat{B}}\hat{N}_\infty | \hat{G}_u) = \hat{E}(I_{\hat{A}}I_{\hat{B}}\hat{N}_\infty | \hat{F}_u) \tag{3.11}$$

Hence,

$$\phi(I_{\hat{B}}\hat{N}_\infty) = \hat{E}(I_{\hat{B}}\hat{N}_\infty) \tag{3.12}$$

Since $(\hat{N}_t)$ is an $(\hat{F}_t,\hat{P})$-martingale, and as $\hat{B} \in \hat{F}_s$,

$$\hat{E}(I_{\hat{B}}\hat{N}_\infty) = \hat{E}(I_{\hat{B}}\hat{N}_s) \tag{3.13}$$

As $\hat{N}_s$ is $\hat{G}_u$-measurable,

$$\phi(I_{\hat{B}}\hat{N}_s) = \hat{E}(I_{\hat{B}}\hat{N}_s) \tag{3.14}$$

By combining (3.12) - (3.14),

$$\phi(I_{\hat{B}}\hat{N}_s) = \phi(I_{\hat{B}}\hat{N}_\infty) \tag{3.15}$$

Let $s > u$. Since $(\hat{N}_t)$ is an $(\hat{F}_t,\hat{P})$-martingale, and as $\hat{B} \in \hat{G}_s \subset \hat{F}_s$,

$$I_{\hat{B}}\hat{N}_s = \hat{E}(\hat{I}_B\hat{N}_\infty|\hat{F}_s) \tag{3.16}$$

whence, as $\hat{G}_u \subset \hat{F}_s$,

$$\hat{E}(I_{\hat{B}}\hat{N}_s|\hat{G}_u) = \hat{E}(\hat{I}_B\hat{N}_\infty|\hat{G}_u) \tag{3.17}$$

As $\tilde{\hat{A}} \cap \hat{B} \in \hat{F}_s$ and since $(\hat{N}_t)$ is a martingale

$$E(I_{\tilde{\hat{A}}}I_{\hat{B}}\hat{N}_\infty|\hat{F}_s) = I_{\tilde{\hat{A}}}I_{\hat{B}}\hat{N}_s \tag{3.18}$$

By combining (3.17)-(3.18), (3.15) also follows for $s > u$. This proves that $(\hat{N}_t)$ is an $(\hat{F}_t,\hat{Q})$-martingale. (qed)

An important special case of the preceding result is when H is of the form

$$H = \{(P(A|H_t)): A \in H_\infty\} \tag{3.19}$$

where (H_t) is a subfiltration of (G_t). In this case, statement (ii) of Proposition 3.1 reads that for any filtration $(\hat{F}_t)$ on any extension $(\hat{\Omega},\hat{F},\hat{P})$ of (Ω,G_∞,P), $(\hat{F}_t) \geq (\hat{G}_t)$, we have

$$(\hat{H}_t) < (\hat{H}_t);\ (\hat{F}_t);\ \hat{P} \Longrightarrow (\hat{G}_t) < (\hat{G}_t);\ (\hat{F}_t);\ \hat{P} \tag{3.20}$$

This result can immediately be applied to solutions of stochastic differential equations. Such an equation is "driven" by a process (Y_t), and together with the equation itself, the specification of

the probability distribution of (Y_t) defines the stochastic differential equation. E.g., the equation

$$X_t = K_t + \int_0^t g_s(X)dZ_s \qquad (3.21),$$

where (Z_t) is a semimartingale and g_t a predictable causal functional (see, e.g., Jacod & Memin (1981) for a specification of the equation and of the concepts involved), is driven by the process $(Y_t) = (Z_t, K_t)$. A <u>regular solution</u> (to be called only "solution" until further notice) of a differential equation is a system $(\Omega, F, P, F_t, X_t, Y_t)$ where (X_t) and (Y_t) are (F_t)-adapted and satisfy the equation, and where (Y_t) is its own cause within (F_t) and has the previously specified distribution.

The application of Proposition 3.1 to the case (3.19) now yields that for the solution $(\Omega, F, P, F_t^{X,Y}, X_t Y_t)$, P is extremal on $F_\infty^{X,Y}$ among the measures Q for which $(\Omega, F_\infty^{X,Y}, Q, F_t^{X,Y}, X_t, Y_t)$ is a solution if and only if every extension $(\hat{\Omega}, \hat{F}, \hat{P})$ of $(\Omega, F_\infty^{X,Y}, P)$ satisfies for every $(\hat{F}_t)$ so that $(\hat{\Omega}, \hat{F}, \hat{P}, \hat{F}_t, \hat{X}_t, \hat{Y}_t)$ is a solution, that $(\hat{X}_t, \hat{Y}_t)$ is its own cause within $(\hat{F}_t)$.

This can also be related to weak uniqueness of the solution, which is that on every solution $(\Omega, F, P, F_t, X_t, Y_t)$, there is no Q on $F_\infty^{X,Y}$, $\neq P$ on $F_\infty^{X,Y}$, so that $(\Omega, F_\infty^{X,Y}, Q, F_t^{X,Y}, X_t, Y_t)$ is a solution. The solution is weakly unique iff the measure is extremal on every solution, hence iff every solution $(\Omega, F, P, F_t, X_t, Y_t)$ satisfies that (X_t, Y_t) is its own casue within F_t relative to P.

There are also other definitions of "solution" of a stochastic differential equation. Like the one we have discussed above, most reduce to a martingale problem, i.e. the problem of determining the set of measures under which a certain set of processes are martingales

(normally also with other conditions on the measures). The main references on these subjects are Jacod (1979), Strook & Varadhan (1979), Yor (1979), Jacod (1980a,b), Strook & Yor (1980), Jacod & Memin (1981) and Lebedev (1983). In most cases, extremality can be used in the same way as above to infer causality.

Extremality of the solution of a martingale problem is an important concept in the theory of martingales. To explain its significance, we need some additional definitions, which will not be used outside this section. For $p \in [1,\infty)$, $H^p(G_t,P)$ is the space of (G_t,P)-martingales (M_t) which satisfy that $\|(M_t)\|^p_{H^p} = E(\sup_t |M_t|^p) < \infty$. A stable subspace of $H^p(G_t,P)$ is a linear space which is closed in the $\|\cdot\|_{H^p}$-topology, and which contains $(I_A M_t)$ and $(M_{t\wedge\tau})$ whenever it contains (M_t) and $A \in G_{-\infty}$ and τ is a (G_t)-stopping time (a T-valued stochastic variable which satisfy $\{\tau \leq t\} \in G_t$ for all t). Note that we here assume that the time axis T is a Borel set.

Assume now that $G_{-\infty}$ is complete, that (G_t) and the elements of H are right continuous, and that $H \subseteq H^1(G_t,P)$. Then a small extension of Theorem 11.2 (p.338) of Jacod (1979) yields that statement (i) of (our) Proposition 3.1 holds if and only if $H^1(G_t,P)$ is the smallest stable subspace of $H^1(G_t,P)$ which contains H and the constant martingale (1). If $H \subseteq H^p(G_t,P)$, then H^1 can be replaced by H^p if H is either finite, or if it consists of continuous martingales only, or if it is of the form (3.19). The two first cases are contained in Corollary 11.4 (p.340) of Jacod (1979), while the last can be found in Mykland (1986).

Stable subspaces are very closely tied to the possibility of representing martingales as stochastic integrals. The relationship can be found in Propositions 4.41 (p.131) and Theorem 4.60 (p.143) in Jacod (1979).

The case $p=2$ is particularly interesting, and we now give an example of its application. Let $\{(M_t^{(1)}), (M_t^{(2)}),\ldots\}$ be a set of orthogonal square integrable (H_t,P)-martingales (i.e. they are elements of $H^2(H_t,P)$, with $(M_t^{(i)}M_t^{(j)})$ as a martingale for $i \neq j$) so that $H^2(H_t,P)$ is the smallest stable subspace of $H^2(H_t,P)$ containing them and (1). Then, (3.20) holds if and only if all square integrable stochastic variables η on G_∞ can be represented

$$\eta = \sum_i \int_{-\infty}^{\infty} f_s^{(i)} dM_s^{(i)} + E(\eta | G_{-\infty}) \qquad (3.22),$$

in the sense of predictable stochastic integrals, as defined in Jacod (1979), ch. II-2, or Elliot (1982), ch. 11.

Representation results like 3.22 are interesting in the theories of filtering and of optimal control of a stochastic process, see e.g., Liptser & Shiryayev (1977), ch. 8-10, and Elliot (1982), ch. 16-18. For econometric work connected to this type of representation, see Harrison & Pliska (1981) and Stricker (1984).

So far, we have been dealing with cases where a conclusion that (G_t) is its own cause is asked for. To end, therefore, note that it is also possible to deal with problems of the form "given a (regular) solution $(\Omega, F_\infty^{X,Y}, P, F_t^{X,Y}, X_t, Y_t)$ of an equation, when does it hold that any extension $(\hat{\Omega}, \hat{F}, \hat{P})$ with filtration $(\hat{F}_t)$, $(\hat{F}_t) \geq (F_t^{\hat{X},\hat{Y}})$, satisfies that $(\hat{X}_t)$ is entirely caused by $(\hat{X}_t, \hat{Y}_t)$ within $(\hat{F}_t)$ relative to $\hat{P}$?". The conditions on the measure P become much less beautiful, but characterizations in times of stability look nicer. E.g., the answer to the question above is that, for any given $p \in (1,\infty)$, this holds iff the smallest stable subspace of $H^p(F_t^{X,Y},P)$ which contains the (F_t^Y)-adapted martingales, also contains the martingales $(E(\eta | F_t^{X,Y}))$, where η describes

$L^p(F^X_\infty,P)$. A further discussion on these matters can be found in Mykland (1984), and related results in Mykland (1986).

4. Application to a stochastic differential equation.

In Sections 4 and 5,

$$T = [0,t_0] \tag{4.1}$$

C^d is the space of all continuous functions $T \to R^d$, $\mathcal{B}_t(C^d)$ is the σ-algebra on C^d making the functions $\varepsilon_u : C^d \to R$,

$$\varepsilon_u(x) = x_u \tag{4.2}$$

measurable for $u \leq t$,

$$\mathcal{B}(C^d) = \mathcal{B}_{t_0}(C^d) \tag{4.3}$$

$$\mathcal{B}_{t+}(C^d) = \bigcap_{u>t} \mathcal{B}_u(C^d) \tag{4.4}$$

A causal functional a_t is a $(\mathcal{B}_{t+}(C^d))$-adapted process on C^d. A (d-dimensional) stochastic process (X_t) which is a continuous (i.e. whose sample-functions are continuous (P-a.s.)) induces a measure μ_X on $\mathcal{B}(C^d)$,

$$\mu_X(B) = P(X(\omega) \in B) \tag{4.5}$$

$X(\omega)$ being the sample-function for given ω. A Wiener process is defined in the usual manner (see, e.g., Lipster & Shiryayev (1977), ch. 4.1) and induces a "Wiener measure" on $\mathcal{B}(C^d)$. Measurability of a process is defined following Lipster & Shiryayev (1977), p.21.

Let (Ω,F,P) be a probability space, F complete, and let $(X_t,F_t)_{t\in T}$ be a continuous d-dimensional stochastic process, F_0 complete. Let μ_X

and μ_W be the measures induced by (X_t) and a d-dimensional Wiener-process respectively. Consider the following statements:

a) There is a (d-dimensional) Wiener-process (W_t, F_t) and a measurable process (α_t, F_t), satisfying

$$X_t = \int_0^t \alpha_s ds + W_t \text{ (P-a.s.) for every } t \in T \tag{4.6}$$

$$\int_0^{t_0} |\alpha_s| ds < \infty \text{ (P-a.s.)} \tag{4.7}$$

$|\cdot|$ being the Euclidean norm.

b)
$$\mu_X \ll \mu_W \tag{4.8}$$

i.e. μ_X is absolutely continuous with respect to μ_W.

b)'
$$\int_0^{t_0} |\alpha_s|^2 < \infty \text{ (P-a.s.)} \tag{4.9}$$

c) (α_t) is of the form (P-a.s.)

$$\alpha_t = a_t(X) \text{ a.e. in } T \tag{4.10}$$

a_t being a causal functional.

c)' (X_t) is its own cause within (F_t).

Liptser & Shiryayev (1977) studies the relationship between statements (a), (b), (b)' and (c) using Girsanov's theorem, see ch. 7.1-7.4 in this book. The picture is completed by statement (c)', as proposition 4.1 will show. First we define some new statements

$$
\begin{aligned}
&(i) && \overset{\Delta}{\Longleftrightarrow} && (b) \ \& \ (c)' \\
&(ii) && \overset{\Delta}{\Longleftrightarrow} && (a) \ \& \ (b) \ \& \ (c)' \\
&(ii)' && \overset{\Delta}{\Longleftrightarrow} && (a) \ \& \ (b)' \ \& \ (c)' \\
&(iii) && \overset{\Delta}{\Longleftrightarrow} && (a) \ \& \ (b) \ \& \ (c) \\
&(iii)' && \overset{\Delta}{\Longleftrightarrow} && (a) \ \& \ (b)' \ \& \ (c)
\end{aligned}
\tag{4.11}.
$$

For example, (iii)' means: "There is a (d-dimensional) Wiener-process (W_t, F_t) and a causal functional a_t satisfying, for every $t \in T$,

$$X_t = \int_0^t a_s(X)ds + W_t \quad (P\text{-a.s.}) \tag{4.12}$$

with

$$\int_0^{t_0} |a_s(X)|^2 ds < \infty \quad (P\text{-a.s.}) \tag{4.13}$$."

Proposition 4.1.

The situation being as described at the beginning of the section, the statements (i) - (iii)' are equivalent. If they apply, the representation (4.6) is unique, i.e. if for every $t \in T$

$$X_t = \int_0^t \beta_s ds + \hat{W}_t \quad (P\text{-a.s.}) \tag{4.14}$$,

$(\hat{W}_t, F_t)$ being a Wiener-process and (β_t, F_t) a measurable process, then

$$P(W_t = \hat{W}_t \text{ for every } t) = 1 \tag{4.15}$$ and

$$P(\alpha_t = \beta_t \text{ for almost every } t) = 1 \tag{4.16}$$.

Proof.

If, for every t,

$$\int_0^t (\beta_s - \alpha_s)ds = W_t - \hat{W}_t \qquad (P\text{-a.s.}) \tag{4.17}$$

the martingale property of $W_t - \hat{W}_t$ implies (4.15) and (4.16), see Lemma 7.1 (p.244) in Liptser & Shiryayev (1977) (this book will in the rest of this proof be known as "LS"). The equivalence between (iii) and (iii)' follows from Theorem 7.5 (p.242) and note 7.2.7 (p.255) in LS, and the implication from (ii)' to (ii) follows from Theorem 7.4 (p.241) in the same book. As (ii) implies (i) trivially, it remains to show that (i) leads to (iii)' and that (iii)' implies (ii)'.

(i) $\Longrightarrow$ (iii)'. Assume first the existence of a (F_t)-adapted (d-dimensional) Wiener-process. In the presence of (4.8), Theorem 7.11 (p.256) in LS guarantees the representation (4.12) for $d = 1$, with (W_t, F_t^X) as a Wiener-process. Going through the same proof, the references to Theorems 5.7 and 6.2 replaced by references to the note at p.170 and to Theorem 6.4 respectively, the same is proved in the multi-dimensional case. (W_t, F_t^X) being a Wiener-process and (F_t^X) being its own cause within (F_t), $(W_t F_t)$ is a Wiener-process, and the implication is proved in this case. In general, we extend the probability space and the framework filtration to include a Wiener-process. As (W_t) in the representation (6.12) is adapted to (F_t^X) and therefore to the original "framework" filtration, the extension can be abandoned after finding (W_t).

(iii)' $\Longrightarrow$ (ii)'. By Ch. 4.7.7 in Liptser & Shiryayev (1977), the solution of (4.12) with side condition (4.13) is weakly unique (actually, the result is only stated for one-dimensional equations, but the generalization to the d-dimensional case is trivial). Hence, since (4.13) must

be satisfied Q-a.s. if $Q \ll P$, P is extremal on F_∞^X among the solutions of (4.12). Hence, by the results of the previous section, (X_t) is its own cause within (F_t).

(qed)

It may be of interest to note that like most other known properties of the equation (4.12), the implication (iii)' $\Longrightarrow$ (ii)' above can be proved with the help of Girsanov's Theorem (see Theorem 6.4 (p.234) in Liptser & Shiryayev (1977)). If $\int_0^{t_0} |a_s(x)|^2 ds$ is bounded in $x \in C^d$, there is a measure $\tilde{P} \sim P$, with $\frac{d\tilde{P}}{dP}$ as F_∞^X-measurable, so that (X_t) is an $(F_t, \tilde{P})$-Wiener process, and hence its own cause within (F_t) relative to $\tilde{P}$. However, it is easily seen that since $\frac{dP}{d\tilde{P}}$ is F_∞^X-measurable, (X_t) is also its own cause within (F_t) relative to P. If we only assume (4.13), we can create a new functional $a_t^{(n)}(x) = a(x) I_{\{t \leq \tau^{(n)}(x)\}}(x)$, where $\tau^{(n)}$ is the (stopping) time when $\int_0^t |a_s(x)|^2 = n$. It is easily seen that the solution $(X_t^{(n)})$ of the equation (4.12) with $a^{(n)}$ instead of a, is (F_t^X)-adapted, and since $(X_t^{(n)})$ converges to (X_t) in probability, the result follows.

5. Criteria of causality.

Given the equation

$$X_t = \int_0^t a_s(X) ds + W_t \qquad (5.1),$$

(X_t) and (W_t) being d-dimensional, we shall in this section find criteria on a_t for the d_1 first components of (X_t) not to be caused (in the (4.8)-sense) by the remaining $d - d_1$ components. This is analogous to similar work done in the framework of Granger-causality,

see in particular Tjøstheim (1981), Sect. 2.1 - 2.2.

In the following, we shall denote $X_t^{(1)}$ the d_1 first components of X_t, and $x_t^{(1)}$, $a_t^{(1)}$, etc., are defined similarly. μ_X and μ_W are the measures induced by X and a Wiener-process respectively.

Proposition 5.1.

Let a probability space (Ω,F,P), a filtration (F_t), a Wiener-process (W_t,F_t) and a (F_t)-adapted continuous process (X_t) governed by (5.1) be given; assume F and F_0 complete. If

$$\mu_X \ll \mu_W \tag{5.2}$$

then $(X_t^{(1)})$ is its own cause within (F_t) if and only if there is a (d-dimensional) causal functional C_t satisfying

$$P\big(a_t^{(1)}(X) = C_t(X^{(1)}) \text{ a.e. in } T\big) = 1 \tag{5.3}$$

If furthermore

$$\mu_X \sim \mu_W \tag{5.4}$$

then $(X_t^{(1)})$ is its own cause iff

$$\mu_W\big(a_t^{(1)}(x) = C_t(x^{(1)}) \text{ a.e. in } T\big) = 1 \tag{5.5}$$

Proof.

This is a corollary to Proposition 4.1 in view of

$$\int_0^{t_0} |a_s^{(1)}(X)|^2 ds \leq \int_0^{t_0} |a_s(X)|^2 ds \tag{5.6}$$

Note the superiority of (5.5) over (5.3). In the latter, the components are allowed to vary independently of one another. In essence, this means that $a_t^{(1)}$ is constant as a function of the $d-d_1$ last arguments.

In the following we shall study a special case where this result can be stated in a stronger form.

(definition)

The metric ρ^d on C^d is given by

$$\forall x,y \in C^d \quad \rho^d(x,y) = \sup_{0 \leq s \leq t_0} |x_s - y_s| \tag{5.7}$$;

the space K^d is defined

$$K^d = \{x : x \in C^d \ \& \ x_0 = 0\} \tag{5.8}$$;

a causal functional a_t is ρ-continuous if, for $x, x_1, \ldots, x_n, \ldots,$

$$a_t(x_n) \longrightarrow a_t(x) \quad \text{a.e. in } T \tag{5.9}$$

when

$$x_n \longrightarrow x \quad \text{uniformly on } T \tag{5.10}$$.

We need the following lemma.

<u>Lemma 5.2.</u>

In the probability space (Ω, F, P), let the d-dimensional Wiener-process (W_t, F_t) be given. Then

$$\forall x \in K^d \quad \forall \varepsilon > 0 \quad P(\omega : \rho^d(x, W(\omega)) < \varepsilon) > 0 \tag{5.11}$$.

Abstract version:

Let μ_W be the measure induced on $\mathcal{B}(C^d)$ by a Wiener-process, let $A \in \mathcal{B}(C^d)$, let

$$\mu_W(A) = 1 \tag{5.12}$$

Then $A \cap K^d$ is dense in K^d under ρ^d.

Proof.

This lemma is easily shown using

$$P(\sup_{0 \leq s \leq t} |W_s| > C) \leq \frac{1}{C^2} \quad \text{for} \quad t \leq 1 \tag{5.13}$$

(this follows from Lemma 4.6 (p.102) in Liptser & Shiryayev (1979) and the technique often used in proving the Heine-Borel theorem (as e.g., in Royden (1968), Theorem 3.15 (p.42).)) (qed)

Proposition 5.3.

Let a probability space (Ω, F, P), a filtration (F_t), a Wiener-process (W_t, F_t) and a (F_t)-adapted process (X_t) governed by (5.1) be given; Assume that F and F_0 are complete, that a_t is ρ-continuous and that

$$\mu_X \sim \mu_W \tag{5.14}$$

$(X_t^{(1)})$ is its own cause within (F_t) if and only if

$$\forall x, y \in K^d \quad x^{(1)} = y^{(1)} \Longrightarrow a^{(1)}(x) = a^{(1)}(y) \tag{5.15}$$

$x^{(1)}$ being the d_1 first components of x, etc.

(This result is a corollary to Prop. 5.1 and Lemma 5.2.)

To end on a less abstract note, consider the causal functional on C^d :

$$a_t(x) = m_0(x(t)) + \int_0^t m_1(x(s))ds + \ldots$$
$$+ \int_0^t ds_1 \ldots \int_0^{s_{n-1}} ds_n m_n(x(s_n)) \tag{5.16}$$

m_i , $i = 0,\ldots,n$, being continuous functions $R^d \to R^d$. a_t is trivially ρ-continuous, and furthermore

$$\int_0^{t_0} |a_s(x)|^2 ds < \infty \tag{5.17}$$

for every $x \in C^d$. By Theorem 7.7 (p.248) in Liptser & Shiryayev (1977), the solution of (5.1) satisfies (5.4). Accordingly, the following result is easily proved.

Proposition 5.4.

Let a probability space (Ω,F,P) , a filtration (F_t) , a Wiener-process (W_t,F_t) and a (F_t)-adapted process (X_t) governed by (5.1) be given; assume that F and F_0 are complete and that a_t is of the form (5.16). Then $(X_t^{(1)})$ is its own cause within (F_t) if and only if the d_1 first components of each of $m_0,\ldots,m_n$ can be expressed as functions of the d_1 first arguments only.

(The proof is carried out in Appendix A.3 of Mykland (1984), where some remarks on possible extentions are made.)

Acknowledgements: The author is grateful to Professor Dag Tjøstheim for valuable advice and comments and to Mrs. Gidske Tresselt for typing the manuscript.

References.

Aalen, Odd O., Ørnulf Borgan, Niels Keiding & Jens Thormann, 1980, Interaction between Life History Events. Nonparametric Analysis for Prospective and Retrospective Data in the Presence of Censoring, Scandinavian Journal of Statistics, 7, 161-171.

Aldous, D.J., 1980. Weak convergence and the general theory of processes, Unpublished manuscript.

Bergstrom, A.R., 1976, ed., Statistical Inference in Continuous Time Economic Models (North Holland).

Beutler, F.J., 1970, Alias-free randomly timed sampling of stochastic processes, IEEE Trans. Information Theory, 16, 147-152.

Brèmaud, P. & M. Yor, 1978. Changes of Filtrations and of Probability Measures, Zeitschrift für Wahrscheinlichkeitstheorie u. verw. Geb., 45, 269-295.

Chamberlain, Gary, 1982. The General Equivalence of Granger and Sims Causality, Econometrica, 50, 569-581.

Chow, Y.S. & Henry Teicher, 1978. Probability Theory (Springer-Verlag, New York).

Elliot, Robert Y., 1982. Stochastic Calculus and Applications (Springer Verlag, Berlin).

Engle, Robert F., David F. Hendry & Jean-Francois Richard, 1983. Exogeneity, Econometrica, 51, 277-304.

Florens, J.P. & Mouchart, M., 1982. A Note on Noncausality, Econometrica, 50, 583-591.

Florens, J.P. & M. Mouchart, 1985. A Linear Theory for Noncausality, Econometrica, 53, 157-175.

Geweke, John, 1982. The Measurement of linear Dependence and Feedback between Multiple Time Series, Journal of the Amer.Stat.ass., 77, 304-313.

Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross spectral methods. Econometrica, 37, 424-438.

Granger, C.W.J., 1980. Testing for Causality. A Personal Viewpoint, Journal of Economic Dynamics and Control, 2, 329-352.

Granger, C.W.J. & M. Hatanaka, 1964. Spectral Analysis of Economic Time Series (Princeton University Press, Princeton NJ).

Granger, C.W.J. & P. Newbold, 1977. Forecasting economic time series (Academic Press, New York).

Harrison, J. Michael, & Stanley R. Pliska, 1981. Martingales and Stochastic Integrals in the Theory of continuous Trading, Stoch.Proc.Appl., 11, 215-260.

Hoover, Douglas N., 1984. Synonymity, generalized martingales and subfiltrations, Ann.Prob., 12, 703-713.

Hoover, Douglas N., & H. Jerome Keisler, 1984. Adapted probability distributions, Transactions of the Am.Mat.Soc., 286, 159-201.

Jacod, Jean, 1979. Calcul Stochastique et Problèmes de Martingales, Lect. Note Math. 714, (Springer Verlag, Berlin).

Jacod, Jean, 1980a. Weak and Strong Solutions of Stochastic Differential Equations, Stochastics, 3, 171-191.

Jacod, Jean, 1980b. Une Condition d'Existence et d'Unicité pour les Solutions Fortes d'Équations Differentielles Stochastiques, Stochastics, 4, 23-38.

Jacod, Jean, & Jean Memin, 1981. Existence of Weak Solutions for Stochastic Differential Equations with Driving Semimartingales, Stochastics, 4, 317-337.

Jeulin, Thiérry, 1980. Semi-Martingales et Grossissement d'une Filtration, Lect. Notes in Math. 833 (Springer Verlag, Berlin).

Jones, Richard H., 1981. In D.F. Finley, ed., Applied Time Series Analysis II (Academic Press).

Lebedev, V.A., 1981. On the Existence of Weak Solutions for Stochastic Differential Equations with Driving Martingales and Random Measures, Stochastics, 9, 37-76.

Liptser, R.S. & A.N. Shiryayev, 1977. Statistics of Random Processes I, (Springer Verlag, New York).

Mouchart, M. & J.M. Rolin, 1979. A Note on Conditional Independence, Séminaire de mathématique appliquée et mécanique, rapport no.129, Institut de mathématique pure et appliquée, Université catholique de Louvain, Louvain.

Mykland, Per A., 1984. Causality in Stochastic Processes (M.S. thesis in mathematics, University of Bergen).

Mykland, Per A., 1986. Stable subspaces over regular solutions of martingale problems, Statistical Report no.15, Dept. of Mathematics, University of Bergen (submitted for publication).

Pierce, D.A. & L.D. Haugh, 1977. Causality in temporal systems: Characterizations and a survey, Journal of Econometrics, 5, 265-293.

Robinson, P.M. 1980. Continuous Model Fitting from Discrete Data, Directions in Time Series, 1980, 263-278.

Royden, H.L. 1968. Real Analysis (Collier Macmillan, New York).

Schweder, T., 1970. Composable Markov Processes, Journal of Applied Probability, 7, 400-410.

Shapiro, H.S. & R.A. Silverman, 1960. Aliasfree sampling of random noise, Journal of SIAM, 8, 141-152.

Sims, C.A. 1972. Money, income and causality, American Economic Review, 62, 540-552.

Stricker, C., 1984. Integral representation in the theory of continuous trading, Stochastics, 13, 249-265.

Strook, D.W., & S.R.S. Varadhan, 1979. Multidimensional Diffusion Processes (Springer-Verlag, New York).

Strook, D.W., & M. Yor, 1980. On extremal solutions of martingale problems, Am.Scient. École Norm. Sup., 13, 95-164.

Suppes, Patrick, 1970. A probabilistic theory of causality, Acta Philosophica Fennica, 24.

Tjøstheim, Dag, 1981. Granger-causality in Multiple Time Series, Journal of Econometrics, 17, 157-176.

Yor, M. 1979. Sur L'Étude des Martingales Continues Extrêmales, Stochastics, 2, 191-196.