Kullback Causality Measures
Author(s): Christian Gourieroux, Alain Monfort and Eric Renault
Source: *Annales d'Économie et de Statistique*, No. 6/7, Modélisation des systèmes dynamiques (Apr. - Sep., 1987), pp. 369-410
Published by: GENES on behalf of ADRES
Stable URL: http://www.jstor.org/stable/20075662
Accessed: 06-02-2018 16:20 UTC

ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 6/7 — 1987

Kullback Causality Measures

Christian GOURIEROUX,
Alain MONFORT, Eric RENAULT *

ABSTRACT. – In this paper we propose causality measures based on the Kullback Information Criterion. These causality measures are applicable in a general context which contains, as special cases, the stationary autoregressive case, considered by GEWEKE, and qualitative models. Estimators of these measures and test procedures are proposed. The nesting of the hypotheses and the asymptotic independence of the test statistics are carefully studied.

Mesures de causalité au sens de Kullback

RÉSUMÉ. – Dans cet article nous proposons des mesures de causalité fondées sur l'information de Kullback. Ces mesures sont utilisables dans un cadre général contenant les modèles autorégressifs, étudiés par GEWEKE, et des modèles qualitatifs. Des estimateurs de ces mesures et des procédures de test sont proposés. L'emboîtement des hypothèses et l'indépendance asymptotique des statistiques de tests sont discutés en détail.

* C. GOURIEROUX : Université Lille-I, Laboratoire de statistiques, B.P. n° 36, 59665 Villeneuve d'Asq, et CEPREMAP; A. MONFORT : Unité de Recherche INSEE, 18, boulevard A. Pinard, 75675 Paris Cedex 14; E. RENAULT : ENSAE, 3, avenue P. Larousse, 92240 Malakoff.

1 Introduction

Following articles by GRANGER [1969] and SIMS [1972], several recent papers deal with definitions and tests of causality (HOSOYA [1977], FLORENS-MOUCHART [1982 and 1985], CHAMBERLAIN [1982], BOUISSOU LAFFONT VUONG [1986]). A paper by GEWEKE [1982] proposes a new direction based on causality measures (see also PIERCE [1979]).

In the present paper we adopt, like GEWEKE, an approach based on causality measures. We show that it is possible to define causality measures in a general context which contains, as special cases, the stationary autoregressive models considered by GEWEKE and qualitative models. These general definitions of causality are based on the Kullback Information Criterion. Measures of the causalities à la GRANGER and à la SIMS are given and their properties are studied. Estimators of these causality measures are proposed and are shown to provide test statistics. The nesting of the various causality hypotheses is carefully considered as well as the asymptotic independence of the test statistics. The quantitative autoregressive case and the qualitative Markov case are studied. In the latter case chi square causality measures are also proposed which, like the Kullback causality measures, can be can decomposed in terms of states and in terms of causality types; the test statistics are also decomposed in a similar way.

In section 2 we define the maintained stochastic framework, which is a Markov process of order p; this process is not assumed to be stationary and allows to consider quantitative or qualitative variables. In section 3 various causality hypotheses are defined in the general framework of section 2. Section 4 shows that, in this general framework, it is possible, by using the Kullback Information Criterion, to propose close forms for the various causality measures; it is also seen that the dependence measure can be decomposed into two unidirectional causality measures and an instantaneous causality measure; moreover the links between Granger causality measures, Sims causality measures and Markov measures are studied; some properties of the causality measures, in particular an invariance property, are also shown. In section 5 the estimation and hypothesis testing problems are treated; in particular two asymptotically equivalent estimators of each causality measure are proposed; the first one is directly deduced from the expression of the theoretical causality measure and the second one is proportional to the likelihood ratio statistic; this second type of estimators provide a natural bridge to hypothesis testing and to the asymptotic independence properties of the test statistics. In section 6 several particular cases are considered: the vector autoregressive models, the homogeneous Markov chains and the non-homogeneous Markov chains. Finally, various technical proofs are gathered in appendices.

2 The Stochastic Framework

We consider a multivariate Markov process of order p; this process is not necessarily homogeneous, we assume that it started at time $t=-p+1$ and it will be denoted by $\{Z_t;\ t \geqslant -p+1\}$. Each component Z_t is partitioned into two subvectors X_t and Y_t whose ranges are respectively $\mathscr{X}$ and $\mathscr{Y}$, some subsets of euclidian spaces; the range of Z_t is $\mathscr{X} \times \mathscr{Y}$. The vectors

$$(X'_t, X'_{t-1}, \ldots, X'_{-p+1})' (Y'_t, Y'_{t-1}, \ldots, Y'_{-p+1})'$$

and $(Z'_t, Z'_{t-1}, \ldots, Z'_{-p+1})$ are denoted by $\mathbf{X}_t$ $\mathbf{Y}_t$ and $\mathbf{Z}_t$ respectively. The probability distribution of the process $\{Z_t;\ t \geqslant -p+1\}$ is defined by:

(i) the probability distribution of the initial vector $\mathbf{Z}_0 = (\mathbf{X}'_0, \mathbf{Y}'_0)'$ which is assumed to have a probability density function, $f_0(\mathbf{z}_0) = f_0(\mathbf{x}_0, \mathbf{y}_0)$, with respect to a product measure $\bigotimes_{i=0}^{p-1} \mu(dz_{-i})$, where $\mu(dz)$ is itself a product measure $\mu_x(dx) \otimes \mu_y(dy)$ on $\mathscr{X} \times \mathscr{Y}$ [these measures could also depend on i];

(ii) the conditional probability distributions of Z_t given $\mathbf{Z}_{t-1}$, for any $t \geqslant 1$, which are assumed to have a p. d. f., $l_{ot}(z_t/\mathbf{z}_{t-1})$, with respect to μ (which, for simplicity, is assumed to be the same for all t); note that the markovian assumption implies that $l_{0t}(z_t/\mathbf{z}_{t-1})$ depends on $\mathbf{z}_{t-1}$ only through $z_{t-1}, \ldots, z_{t-p}$ and this p. d. f. is also denoted by $l_{0t}(z_t/z_{t-1}^{t-p})$ or $l_{0t}(x_t, y_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1})$ or $l_{0t}(x_t, y_t/x_{t-1}^{t-p}, y_{t-1}^{t-p})$ where $z_{t-1}^{t-p} = (z'_{t-1}, \ldots, z'_{t-p})$ and x_{t-1}^{t-p}, y_{t-1}^{t-p} have similar meanings.

The joint p. d. f. of $\mathbf{Z}_t = (Z'_t, Z'_{t-1}, \ldots, Z'_{-p+1})'$, with respect to $\bigotimes_{i=-p+1}^{t} \mu(dz_i)$ is:

$$l_{0t}(\mathbf{z}_t) = f_0(\mathbf{z}_0) \cdot \prod_{i=1}^{t} (l_{0i}(z_i/\mathbf{z}_{i-1})$$

$$= f_0(z_0) \prod_{i=1}^{t} l_{01}(z_i/z_{i-1}^{i-p})$$

Vatious marginal or conditional p. d. f., can now be derived. The marginal p. d. f. of $\mathbf{X}_t$, with respect to $\bigotimes_{i=-p+1}^{t} \mu_x(dx_i)$ is

$$l_{0t}(\mathbf{x}_t) = \int_{\mathscr{Y}^{p+1}} l_{0t}(\mathbf{z}_t) \bigotimes_{i=-p+1}^{t} \mu_y(dy_i)$$

The conditional p. d. f. of X_t given $\mathbf{X}_{t-1}$, $\mathbf{Y}_{t-1}$ with respect to μ_x is

$$l_{0t}(x_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1}) = \int_{\mathscr{Y}} l_{0t}(x_t, y_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1})\, \mu_y(dy_t)$$

where $\mathbf{x}_{t-1}$, $\mathbf{y}_{t-1}$ can be replaced by x_{t-1}^{t-p}, y_{t-1}^{t-p}.

Similarly we have:

$$l_{0t}(\mathbf{y}_t) = \int_{\mathscr{X}^{p+t}} l_{0t}(\mathbf{z}_t) \bigotimes_{i=-p+1}^{t} \mu_x(dx_i),$$

$$l_{0t}(y_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1}) = \int_{\mathscr{X}} l_{0t}(x_t, y_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1})\, \mu_x(dx_t),$$

where $\mathbf{x}_{t-1}$, $\mathbf{y}_{t-1}$ can be replaced by $x_{t-1}^{t-p}, y_{t-1}^{t-p}$.

The conditional p. d. f. of X_t given $\mathbf{X}_{t-1}$, with respect to $\mu_x(dx_t)$ is:

$$l_{0t}(x_t/\mathbf{x}_{t-1}) = \frac{l_{0t}(\mathbf{x}_t)}{l_{0t}(\mathbf{x}_{t-1})}.$$

Note that, in general, this p. d. f. depends on all the past values $(x_{t-1}, \ldots, x_{-p+1})$; in other words the process $\{X_t;\ t \geqslant -p+1,\}$ is not Markovian. Similar remarks hold for $\{Y_t;\ t \geqslant -p+1\}$.

3 Causality Hypotheses

We now assume that the process $\{Z_t;\ t \geqslant -p+1\}$ has been observed for $t = -p+1, \ldots, T$ and we define various hypotheses on the joint distribution of $\mathbf{Z}_T$; in particular, we consider non causality hypotheses based on GRANGER' [1969] definitions.

3.*a.* Basic Hypotheses

i) The General Hypothesis H is simply made of the probability distributions of $\mathbf{Z}_T$ satisfying the assumptions of the previous section, i. e. the markovian property and the existence of p. d. f.'s with respect to product measures; the joint p. d. f. of $\mathbf{Z}_T$ is denoted by $l_{0t}(\mathbf{z}_T)$.

ii) ***Non Causality from*** X ***to*** Y : H_1

This hypothesis is made of the probability distributions of H such that Y_t and $\mathbf{X}_{t-1}$ are conditionally independent given $\mathbf{Y}_{t-1}$, for any $t = 1, \ldots, T$. This hypothesis can be written:

$$l_{0t}(y_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1}) = l_{0t}(y_t/\mathbf{y}_{t-1}), \qquad 1 \leqslant t \leqslant T$$

iii) ***Non Causality from*** Y ***to*** X : H_2

It is the symmetrical hypothesis of conditional independence of X_t and $\mathbf{Y}_{t-1}$ given $\mathbf{X}_{t-1}$ for any $t = 1, \ldots, T$:

$$l_{0t}(x_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1}) = l_{0t}(x_t/\mathbf{x}_{t-1})$$

iv) ***Non Instantaneous Causality:*** H_3

This hypothesis is made of the probability distributions of H such that X_t and Y_t are conditionally independent given $\mathbf{X}_{t-1}$, and $\mathbf{Y}_{t-1}$ for any $t=1, \ldots, T$; this hypothesis can be written in three equivalent forms:

$$l_{0t}(x_t/\mathbf{x}_{t-1}, \mathbf{y}_t) = l_{0t}(x_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1})$$

or

$$l_{0t}(y_t/\mathbf{x}_t, y_{t-1}) = l_{0t}(y_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1})$$

or

$$l_{0t}(x_t, y_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1}) = l_{0t}(x_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1})\, l_{0t}(y_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1})$$
$$1 \leqslant t \leqslant T.$$

3.*b.* Cross Hypotheses

From the previous basic hypotheses it is possible to define four other hypotheses obtained as intersections of the basic hypotheses.

global non causality from X to Y : $H_{13} = H_1 \cap H_3$;

global non causality from Y to X : $H_{23} = H_2 \cap H_3$;

bidirectional non causality : $H_{12} = H_1 \cap H_2$;

independence : $H_{123} = H_1 \cap H_2 \cap H_3$.

Note that, under H_{123}, the p. d. f. of $\mathbf{Z}_r$ can be written:

$$l_{0T}(z_T) = f_0(\mathbf{z}_0) \prod_{t=1}^{T} l_{0t}(x_t, y_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1})$$

or, using H_3 :

$$l_{0T}(z_T) = f_0(\mathbf{z}_0) \prod_{t=1}^{T} l_{0t}(x_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1}) \prod_{t=1}^{T} l_{0t}(y_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1})$$

or, using H_1 and H_2:

$$l_{0T}(z_T) = f_0(\mathbf{z}_0) \prod_{t=1}^{T} l_{0t}(x_t/\mathbf{x}_{t-1}) \prod_{t=1}^{T} l_{0t}(y_t/\mathbf{y}_{t-1})$$

Therefore, H_{123} is the conditional independence of $X_t^1 = (X_t', \ldots, X_1')'$ and $Y_t^1 = (Y_t', \ldots, Y_1')'$ given $\mathbf{Z}_0 = (Z_0', \ldots, Z_{-p+1}')'$.

3.*c.* How to Nest These Hypotheses?

The eight hypotheses H, H_1, H_2, H_3, H_{13}, H_{23}, H_{12}, H_{123} can be nested in six ways corresponding to the six possible orderings of the three numbers 1, 2, 3:

TABLE 1

Ordering	Nesting sequence
123.	$H \supset H_1 \supset H_{12} \supset H_{123}$
132.	$H \supset H_1 \supset H_{13} \supset H_{123}$
213.	$H \supset H_2 \supset H_{12} \supset H_{123}$
231.	$H \supset H_2 \supset H_{23} \supset H_{123}$
312.	$H \supset H_3 \supset H_{13} \supset H_{123}$
321.	$H \supset H_3 \supset H_{23} \supset H_{123}$

4 Kullback Causality Measures

4.*a.* The Principle

The unknown true probability distribution of $\mathbf{Z}_T$, $l_{0T}(\mathbf{z}_T)$, is assumed to belong to the maintained hypothesis H and, using the Kullback Information Criterion, it is possible to define a "discrepancy" between l_{0T} and any of the hypotheses defined above and denoted by $\tilde{H}$:

$$D(\tilde{H}/l_{0T}) = \frac{1}{T} \underset{l \in \tilde{H}}{\mathrm{Min}}\, E_0 \operatorname{Log} \frac{l_{0T}(\mathbf{Z}_T)}{l(\mathbf{Z}_T)}$$

$$= \frac{1}{T} \underset{l \in \tilde{H}}{\mathrm{Min}} \int \operatorname{Log} \frac{l_{0T}(\mathbf{z}_T)}{l(\mathbf{z}_T)} l_{0T}(\mathbf{z}_T) \bigotimes_{t=-p+1}^{T} \mu(dz_t)$$

A p. d. f. in $\tilde{H}$ for which the minimum is reached is called *a pseudo-true p. d. f.*

The previous definition provides seven discrepancies indexed in the same way as the hypotheses: D_1, D_2, D_3, D_{12}, D_{13}, D_{23}, D_{123}.

Since these measure are obtained from a minimisation on a subset of H it is clear that these measures increase for any the six sequences of nested hypotheses presented in Table 1. For instance, for the first sequence of Table 1, we have:

$$0 \leqslant D_1 \leqslant D_{12} \leqslant D_{123}.$$

From these discrepancies it is possible to define causality measures. Let us consider, for instance, the causality from X to Y; this causality can be measured in different ways according to the hypothesis which is considered as maintained; if the maintained hypothesis is H, the causality from X to Y can be measured as the difference between the discrepancy associated with H_1 and the discrepancy associated with H i. e. $D_1 - 0 = D_1$; if the maintained hypothesis is H_2 (non causality from Y to X) the causality of X on Y can be measured by $D_{12} - D_2$; similarly we obtain $D_{123} - D_{23}$ if

H_{23} is maintained, $D_{13}-D_3$ if H_3 is maintained. In other words, with each nesting sequence presented in Table 1, it is possible to associate three causality measures related to the causality from X to Y, the causality from Y to X and the instantaneous causality. These measures are summarised in Table 2.

TABLE 2

Ordering	Causality from X to Y	Causality from Y to X	Instantaneous causality	Total (Dependence)
123.	D_1	$D_{12}-D_1$	$D_{123}-D_{12}$	D_{123}
132.	D_1	$D_{123}-D_{13}$	$D_{13}-D_1$	D_{123}
213.	$D_{12}-D_2$	D_2	$D_{123}-D_{12}$	D_{123}
231.	$D_{123}-D_{23}$	D_2	$D_{23}-D_2$	D_{123}
312.	$D_{13}-D_3$	$D_{123}-D_{13}$	D_3	D_{123}
321.	$D_{123}-D_{23}$	$D_{23}-D_3$	D_3	D_{123}

From Table 2 it is clear that, for each of the three types of causality, there are four causality measures; for instance, the four possible measures for causality from X to Y are D_1, $D_{12}-D_2$, $D_{13}-D_3$ and $D_{123}-D_{23}$. It also appears that, to each ordering, corresponds a decomposition of the dependence measure D_{123}, into three causality measures.

This lack of uniqueness of the causality measures is not satisfactory; however we shall see that, for each type of causality, three of the four possible measures are equal.

4.*b.* Derivation of the Discrepancies

It is possible to derive a close expression for six of the seven discrepancies; the only one for which a problem arises is D_{12}.

Let us, for instance, consider D_{13}. We have:

$$D_{13}=\frac{1}{T}\underset{l\in H_{13}}{\text{Min}}\ E_0 \text{ Log } \frac{l_{0T}(\mathbf{Z}_T)}{l(\mathbf{Z}_T)}.$$

In H_{13} the p. d. f. s have the following form:

$$l(z_T)=f(\mathbf{x}_0,\mathbf{y}_0)\prod_{t=1}^{T} l_t(x_t/\mathbf{x}_{t-1},\mathbf{y}_{t-1})\, l_t(y_t/\mathbf{y}_{t-1})$$

Moreover, from the lemma of appendix 0, it follows that under H_{13} the process Y_t is Markovian of order p and therefore.

$$D_{13} = \frac{1}{T} \underset{l \in H_{13}}{\text{Min}} \ E_0 \text{ Log } \frac{f_0(X_0, Y_0) \prod_{t=1}^{T} l_{0T}(X_t, Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{f(X_0, Y_0) \prod_{t=1}^{T} l_t(X_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p}) \, l_t(Y_t/Y_{t-1}^{t-p})}$$

$$= \frac{1}{T} \left\{ \text{Min } E_0 \text{ Log } \frac{f_0(X_0, Y_0)}{f(X_0, Y_0)} + \text{Min} \sum_{t=1}^{T} E_0 \text{ Log } \frac{l_{0t}(X_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_t(X_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})} \right.$$

$$\left. + \text{Min} \sum_{t=1}^{Y} E_0 \frac{l_{0t}(Y_t/X_t^{t-p}, Y_{t-1}^{t-p})}{l_t(Y_t/Y_{t-1}^{t-p})} \right\}$$

It is clear that the first two terms are equal to zero since it is possible to take the denominator equal to the numerator. Let us now consider the third term. The optimisation problem appearing in this term can be written:

$$\text{Max} \sum_{t=1}^{T} E_0 \text{ Log } l_t(Y_t/Y_{t-1}^{t-p})$$

or

$$\text{Max} \sum_{t=1}^{T} E_0 E_0 \left[\text{Log } l_t(Y_t/Y_{t-1}^{t-p})/Y_{t-1}^{t-p}\right]$$

It follows from the Kullback inequality that, for each t and for a given Y_{t-1}^{t-p}, the conditional expectation is maximised by choosing $l_t(Y_t/Y_{t-1}^{t-p}) = l_{0t}(Y_t/Y_{t-1}^{t-p})$. This result shows that the discrepancy D_{13} is given by

$$D_{13} = \frac{1}{T} \sum_{t=1}^{T} E_0 \text{ Log } \frac{l_{0t}(Y_t/X_t^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})}$$

The pseudo-true p. d. f. is

$$l_0^*(\mathbf{z}_T) = f_0(\mathbf{x}_0, \mathbf{y}_0) \prod_{t=1}^{T} l_{0t}(x_t/x_{t-1}^{t-p}, y_{t-1}^{t-p}) \, l_{0t}(y_t/y_{t-1}^{t-p}).$$

It is also worth noting that if the general hypothesis H is replaced by the more restrictive hypothesis H* in which the homogeneity of the Markov process is imposed and the initial p. d. f. is the invariant p. d. f., then the process Z_t is stationary and the p. d. f. s in H_{13} do not depend on t; in this case D_{13} becomes:

$$D_{13}^* = E_0 \text{ Log } \frac{l_0(Y_t/X_t^{t-p}, Y_{t-1}^{t-p})}{l_0(Y_t/Y_{t-1}^{t-p})}.$$

Also note that if H is replaced by H** in which only the additional assumption of homogeneity is imposed, $l_{0t}(Y_t/X_t^{t-p}, Y_{t-1}^{t-p})$ is time invariant whereas, in general, $l_{0t}(Y_t/Y_{t-1}^{t-p})$ is time dependent. Therefore D_{13} becomes:

$$D_{13}=E_0\ l_0(Y_t/X_t^{t-p},\ Y_{t-1}^{t-p})-\frac{1}{T}\sum_{t=1}^{T} E_0 \text{ Log } l_{0t}(Y_t/Y_{t-1}^{t-p}).$$

However the distribution whose p. d. f. is $l_{0t}(Y_t/Y_{t-1}^{t-p})$ converges, as t goes to infinity, to a distribution corresponding to the stationary regime whose p. d. f. is $l_0(Y_t/Y_{t-1}^{t-p})$; then E_0 Log $l_{0t}(Y_t/Y_{t-1}^{t-p})$ converges to E_0 Log $l_0(Y_t/Y_{t-1}^{t-p})$ and it follows that the Cesaro sum D_{13} converges to D_{13}^*.

The forms of the other discrepancies, except D_{12}, can be derived in the same way (see appendix 1) and the results are gathered in the following theorem.

THEOREM 1 : Under the general hypothesis H, we have:

$$D_1 \ =\frac{1}{T}\sum_{t=1}^{T} E_0 \text{ Log } \frac{l_{0t}(Y_t/X_{t-1}^{t-p},\ Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})}$$

$$D_2 \ =\frac{1}{T}\sum_{t=1}^{T} E_0 \text{ Log } \frac{l_{0t}(X_t/X_{t-1}^{t-p},\ Y_{t-1}^{t-p})}{l_{0t}(X_t/X_{t-1}^{t-p})}$$

$$D_3 \ =\frac{1}{T}\sum_{t=1}^{T} E_0 \text{ Log } \frac{l_{0t}(X_t,\ Y_t/X_{t-1}^{t-p},\ Y_{t-1}^{t-p})}{l_{0t}(X_t/X_{t-1}^{t-p},\ Y_{t-1}^{t-p})\, l_{0t}(Y_t/X_{t-1}^{t-p},\ Y_{t-1}^{t-p})}$$

$$D_{13} =\frac{1}{T}\sum_{t=1}^{T} E_0 \text{ Log } \frac{l_{0t}(Y_t/X_t^{t-p},\ Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})}$$

$$D_{23} =\frac{1}{T}\sum_{t=1}^{T} E_0 \text{ Log } \frac{l_{0t}(X_t/X_{t-1}^{t-p},\ Y_t^{t-p})}{l_{0t}(X_t/X_{t-1}^{t-p})}$$

$$D_{123} =\frac{1}{T}\sum_{t=1}^{T} E_0 \text{ Log } \frac{l_{0t}(X_t,\ Y_t/X_{t-1}^{t-p},\ Y_{t-1}^{t-p})}{l_{0t}(X_t/X_{t-1}^{t-p})\, l_{0t}(Y_t/Y_{t-1}^{t-p})}$$

Under H*, i. e. under the additional assumptions of homogeneity and stationarity, these measures do not depend on T and are equal to D_1^*, D_2^*, D_3^*, D_{13}^*, D_{123}^* obtained from D_1, D_2, D_3, D_{13}, D_{23}, D_{123} respectively by simply taking any term of the Cesaro sum. Under H**, i. e. the additional assumption of homogeneity only, the previous results remain exactly valid for D_3 and asymptotically valid for the other measures.

From classical results on Kullback Information Criterion we know that all these discrepancies are non negative (see appendix 2) and that they are equal to zero if, and only if, the corresponding hypothesis is satisfied.

4.*c*. Derivation of the Causality Measures

In subsection 4.*a* we have seen that for each of the three types of causality there are four possible measures. In this subsection we prove that three among these four measures are equal, mamely the ones in which D_{12} does not appear. Note that D_{12} never appears in the nesting sequences associated with the orderings 132, 231, 312, or 321 that is to say the orderings in which 3 is not in the last position.

THEOREM 2 (causality from X to Y) :

$$D_1 = D_{13} - D_3 = D_{123} - D_{23}.$$

Proof:

$$D_{13} - D_3 = \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log}\left[\frac{l_{0t}(Y_t/X_t^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})} \times \frac{l_{0t}(X_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})\, l_{0t}(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(X_t, Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}\right]$$

$$= \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})}$$

$$= D_1$$

$$D_{123} - D_{23} = \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(X_t, Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(X_t/X_{t-1}^{t-p})\, l_{0t}(Y_t/Y_{t-1}^{t-p})} \frac{l_{0t}(X_t/X_{t-1}^{t-p})}{l_{0t}(X_t/X_{t-1}^{t-p}, Y^{t-p})}$$

$$= \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(X_t, Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})\, l_{0t}(X_t/X_{t-1}^{t-p}, Y_t^{t-p})}$$

$$= \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})}$$

$$= D_1 \qquad \square$$

Since three possible measures are equal and, among them, the more natural one, i. e. D_1, this leads to the following definition.

DEFINITION 1 : The Kullback measure of causality from × and Y is:

$$C_{X \to Y} = D_1 = D_{13} - D_3 = D_{123} - D_{23}$$

$$= \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})}$$

$$= \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(X_t, Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})\, l_{0t}(X_t/X_{t-1}^{t-p}, Y_t^{t-p})}$$

Under H* (stationarity) or H** (homogeneity) this measure is defined by:

$$C_{X \to Y} = D_1^* = D_{13}^* - D_3^* = D_{123}^* - D_{23}^*$$

$$= E_0 \operatorname{Log} \frac{l_0(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_0(Y_t/Y_{t-1}^{t-p})}$$

$$= E_0 \operatorname{Log} \frac{l_0(X_t, Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_0(Y_t/Y_{t-1}^{t-p})\, l_0(X_t/X_{t-1}^{t-p}, Y_t^{t-p})}$$

where l_0 without the index t denotes a stationary p. d. f.

THEOREM 3 (causality from Y to X) :

$$D_2 = D_{23} - D_3 = D_{123} - D_{13}$$

Proof: Deduced from that of theorem 2 by symmetry. □

DEFINITION 2 : The Kullback measure of causality from Y to X is:

$$C_{Y \to X} = D_2 = D_{23} - D_3 = D_{123} - D_{13}$$

$$= \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(X_t / X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(X_t / X_{t-1}^{t-p})}$$

$$= \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(X_t, Y_t / X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(X_t / X_{t-1}^{t-p})\, l_{0t}(Y_t / X_t^{t-p}, Y_{t-1}^{t-p})}$$

The same simplifications as in definition 1 hold under H* and H**

Since, by definition, the sum of the three causality measures is always D_{123}, the dependence measure (see Table 2), we immediately see that $D_3 = D_{13} - D_1 = D_{23} - D_2$ and this leads to the following definition.

DEFINITION 3 : The Kullback measures of instantaneous causality is:

$$C_{X \leftrightarrow Y} = D_3 = D_{13} - D_1 = D_{23} - D_2$$

$$= \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(X_t, Y_t / X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(X_t / X_{t-1}^{t-p}, Y_{t-1}^{t-p})\, l_{0t}(Y_t / X_{t-1}^{t-p}, Y_{t-1}^{t-p})}$$

$$= \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(X_t / X_{t-1}^{t-p}, Y_t^{t-p})}{l_{0t}(X_t / X_{t-1}^{t-p}, Y_{t-1}^{t-p})}$$

$$= \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(Y_t / X_t^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t / X_{t-1}^{t-p}, Y_{t-1}^{t-p})}$$

Under H* or H** the same simplifications as in definitions 1 and 2 hold.

DEFINITION 4 : The Kullback measure of dependence is:

$$C_{X, Y} = D_{123}$$

$$= \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(X_t, Y_t / X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(X_t / X_{t-1}^{t-p})\, l_{0t}(Y_t / Y_{t-1}^{t-p})}$$

with the usual simplifications under H* and H**. The global causality measure from X to Y is defined by:

$$C_{X \Rightarrow Y} = C_{X \to Y} + C_{X \leftrightarrow Y}.$$

From the previous results we obviously have a decomposition generalising that proposed by GEWEKE [1982]:

$$\begin{aligned} C_{X,Y} &= C_{X \leftrightarrow Y} + C_{X \to Y} + C_{Y \to X} \\ &= C_{X \Rightarrow Y} + C_{Y \to X} \\ &= C_{Y \Rightarrow X} + C_{X \to X} \end{aligned}$$

It is worth noting that, in a general framework, measures based on difference between discrepancies will depend on the nesting sequence chosen. It is remarkable to find that, in the present case, these measures are identical for four nesting sequences: 132, 231, 312 and 321. For the sequences in which 3 is in the last position, that is the sequences in which the instantaneous causality is considered in the last position, the result no longer holds in general. This seems to suggest that, when studying the links between two processes, one ought to consider first the whole influence of one process on the other (i. e. the exogeneity properties).

Also note that the previous decomposition is valid without the markovian assumption; however the interpretations in terms of causality obviously rest on the markovian assumption.

4.*d.* Alternative Causality Measures

In this subsection we derive causality measures based on Sims'approach to causality and we discuss their links with the previous ones, based on Granger's approach.

DEFINITION 5 : The Kullback measure of causality from X to Y, in Sims'-sense, is:

$$CS_{X \to Y} = \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(X_t / X_{t-1}^{t-p}, Y_T^{t-p})}{l_{0t}(X_t / X_{t-1}^{t-p}, Y_t^{t-p})}.$$

The Kullback measure of global causality from X to Y in Sims'sense is

$$CS_{X \Rightarrow Y} = \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(X_t / X_{t-1}^{t-p}, Y_T^{t-p})}{l_{0t}(X_t / X_{t-1}^{t-p}, Y_{t-1}^{t-p})}$$

In order to establish the links between $CS_{X \to Y}$ and $C_{X \to Y}$ we also need a measure of the non markovian character of process Y, called Markov measure.

DEFINITION 6 : The Markov measure of Y is

$$M_Y = \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(Y_t / X_0^{1-p}, Y_{t-1}^{1-p})}{l_{0t}(Y_t / Y_{t-1}^{t-p})}$$

THEOREM 4:

$$C_{X\to Y} = CS_{X\to Y} + M_Y$$
$$C_{X\Rightarrow Y} = CS_{X\Rightarrow Y} + M_Y$$

(symmetrical equalities obviously hold).

Proof: From definition 1 we have:

$$C_{X\to Y} = \frac{1}{T}\sum_{t=1}^{T} E_0 \operatorname{Log}\frac{l_{0t}(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})}$$

$$= \frac{1}{T}\sum_{t=1}^{T} E_0 \operatorname{Log}\frac{l_{0t}(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})}\,\frac{l_{0t}(X_t/X_{t-1}^{t-p}, Y_t^{t-p})}{l_{0t}(X_t/X_{t-1}^{t-p}, Y_t^{t-p})}$$

$$= \frac{1}{T} E_0 \operatorname{Log}\frac{\prod_{t=1}^{T} l_{0t}(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})\, l_{0t}(X_t/X_{t-1}^{t-p}, Y_t^{t-p})}{\prod_{t=1}^{T} l_{0t}(Y_t/Y_{t-1}^{t-p}) \prod_{t=1}^{T} l_{0t}(X_t/X_{t-1}^{t-p}, Y_t^{t-p})}$$

$$= \frac{1}{T} E_0 \operatorname{Log}\frac{l_{0T}(X_T^1, Y_T^1/X_0^{1-p}, Y_0^{1-p})}{\prod_{t=1}^{T} l_{0t}(Y_t/Y_{t-1}^{t-p}) \prod_{t=1}^{T} l_{0t}(X_t/X_{t-1}^{t-p}, Y_t^{t-p})}$$

$$= \frac{1}{T} E_0 \operatorname{Log}\frac{l_{0T}(X_T^1/Y_T^1, X_0^{1-p}, Y_0^{1-p})\, l_{0T}(Y_T^1/X_0^{1-p}, Y_0^{1-p})}{\prod_{t=1}^{T} l_{0t}(Y_t/Y_{t-1}^{t-p}) \prod_{t=1}^{T} l_{0t}(X_t/X_{t-1}^{t-p}, Y_t^{t-p})}$$

$$= \frac{1}{T}\sum_{t=1}^{T} E_0 \operatorname{Log}\frac{l_{0t}(X_t/X_{t-1}^{t-p}, Y_T^{1-p})}{l_{0t}(X_t/X_{t-1}^{1-p}, Y_t^{1-p})}$$
$$+ \frac{1}{T}\sum_{t=1}^{T} E_0 \operatorname{Log}\frac{l_{0t}(Y_t/X_0^{1-p}, Y_{t-1}^{1-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})}.$$

Using the Markovian property, $1-p$ can be replaced by $t-p$ in the first sum and we get:

$$C_{X\to Y} = CS_{X\to Y} + M_Y.$$

Moreover from definition 5 we have:

$$CS_{X\Rightarrow Y} - CS_{X\to Y} = \frac{1}{T}\sum_{t=1}^{T} E_0 \operatorname{Log}\frac{l_{0t}(X_t/X_{t-1}^{t-p}, Y_t^{t-p})}{l_{0t}(X_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}$$
$$= C_{X\leftrightarrow Y} \text{ (from definition 3)}$$

Therefore we have:

$$CS_{X\leftrightarrow Y} = CS_{X\to Y} + C_{X\leftrightarrow Y}$$

and adding M_Y to both sides

$$CS_{X\Rightarrow Y} + M_Y = C_{X\to Y} + C_{X\leftrightarrow Y}$$
$$= C_{X\Rightarrow Y} \qquad \square$$

Let us now assume that H* holds, i.e. that the process Z is homogenous and stationary. Let us also add the technical assumption

$$\text{(A)} \qquad \mathrm{M_Y} = \frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \mathrm{E_0 \, Log} \frac{l_{0t}(\mathrm{Y}_t/\mathrm{X}_0^{1-p}, \mathrm{Y}_{t-1}^{1-p})}{l_{0t}(\mathrm{Y}_t/\mathrm{Y}_{t-1}^{t-p})}$$

converges, when T goes to infinity, to a limit denoted by M_Y^*, or $\mathrm{E_0 \, Log} \dfrac{l_0(\mathrm{Y}_t/\mathrm{Y}_{t-1}^{-\infty})}{l_0(\mathrm{Y}_t/\mathrm{Y}_{t-1}^{t-p})}$, we have the following theorem.

THEOREM 5 : Under H* and (A), $\mathrm{CS}_{\mathrm{X}\to\mathrm{Y}}$ and $\mathrm{CS}_{\mathrm{X}\Rightarrow\mathrm{Y}}$ converge, as $\mathrm{T}\to\infty$, to limits denoted respectively by

$$\mathrm{CS}^*_{\mathrm{X}\to\mathrm{Y}} \quad \text{and} \quad \mathrm{CS}^*_{\mathrm{X}\Rightarrow\mathrm{Y}},$$

or

$$\mathrm{E_0 \, Log} \frac{l_0(\mathrm{X}_t/\mathrm{X}_{t-1}^{t-p}, \mathrm{Y}_{\infty}^{t-p})}{l_0(\mathrm{X}_t/\mathrm{X}_{t-1}^{t-p}, \mathrm{Y}_{t}^{t-p})} \quad \text{and} \quad \mathrm{E_0 \, Log} \frac{l_0(\mathrm{X}_t/\mathrm{X}_{t-1}^{t-p}, \mathrm{Y}_{\infty}^{t-p})}{l_0(\mathrm{X}_t/\mathrm{X}_{t-1}^{t-p}, \mathrm{Y}_{t-1}^{t-p})}.$$

The following equalities hold:

$$\mathrm{C}_{\mathrm{X}\to\mathrm{Y}} = \mathrm{CS}^*_{\mathrm{X}\to\mathrm{Y}} + \mathrm{M}^*_\mathrm{Y}$$
$$\mathrm{C}_{\mathrm{X}\Rightarrow\mathrm{Y}} = \mathrm{CS}^*_{\mathrm{X}\Rightarrow\mathrm{Y}} + \mathrm{M}^*_\mathrm{Y}.$$

Moreover each term of the Cesaro sums, appearing in $\mathrm{CS}_{\mathrm{X}\to\mathrm{Y}}$ and $\mathrm{CS}_{\mathrm{X}\Rightarrow\mathrm{Y}}$ converge respectively to $\mathrm{CS}^*_{\mathrm{X}\to\mathrm{Y}}$ and $\mathrm{CS}^*_{\mathrm{X}\Rightarrow\mathrm{Y}}$ when T goes to infinity.

Proof: See appendix 3.

4.e. Some Properties of the Causality Measures

A first property which is worth mentioning is the invariance of the various causality measures or Markov measures introduced above with respect to one to one transformations of the X_t's and the Y_s's.

THEOREM 6 : Let $(\tilde{\mathrm{X}}_t, \tilde{\mathrm{Y}}_t)$ be the process defined by $\tilde{\mathrm{X}}_t = g(\mathrm{X}_t)$ and $\tilde{\mathrm{Y}}_t = h(\mathrm{Y}_t)$, where g and h are one to one transformations. All the causality or Markov measures defined for the process $(\tilde{\mathrm{X}}_t, \tilde{\mathrm{Y}}_t)$ are identical to the corresponding ones for $(\mathrm{X}_t, \mathrm{Y}_t)$.

Proof: All the measures considered above are based on the Kullback information criterion and, therefore, they are invariant with respect to a change of the dominating measure $\mu = \mu_x \otimes \mu_y$. So, for instance, the conditional probability distributions of $\tilde{\mathrm{Y}}_t$ given $\tilde{\mathrm{Y}}_{t-1}^{t-p}$ and of $\tilde{\mathrm{Y}}_t$ given $\tilde{\mathrm{X}}_{t-1}^{t-p}$, $\tilde{\mathrm{Y}}_{t-1}^{t-p}$ have density functions with respected to μ_2^h, the image of μ_2 by h, which are respectively

$$l_{0t}[h^{-1}(\tilde{y}_t)/h^{-1}(\tilde{y}_{t-1}), \ldots, h^{-1}(\tilde{y}_{t-p})]$$

and

$$l_{0t}[h^{-1}(\tilde{y}_t)/g^{-1}(\tilde{x}_{t-1}), \ldots, g^{-1}(\tilde{x}_{t-p}); h^{-1}(\tilde{y}_{t-1}), \ldots, h^{-1}(\tilde{y}_{t-p})];$$

Therefore the causality measure from $\tilde{X}$ to $\tilde{Y}$ is

$$C_{\tilde{X}\to\tilde{Y}}=$$

$$\frac{1}{T}\sum_{t=1}^{T} E_0 \operatorname{Log}\frac{l_{0t}[h^{-1}(\tilde{Y}_t)/g^{-1}(\tilde{X}_{t-1}), ., g^{-1}(\tilde{X}_{t-p}); h^{-1}(\tilde{Y}_{t-1}), ., h^{-1}(\tilde{Y}_{t-p})]}{l_{0t}[h^{-1}(\tilde{Y}_t)/h^{-1}(\tilde{Y}_{t-1}), \ldots, h^{-1}(\tilde{Y}_{t-p})]}$$

$$=\frac{1}{T}\sum_{t=1}^{T} E_0 \operatorname{Log}\frac{l_{0t}(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})}=C_{X\to Y} \qquad \square$$

It is also interesting to consider the behaviour of the various measures when X_t and Y_t are partitioned into independent subvectors.

THEOREM 7: Let us consider the partitions $X_t=(X'_{1t}, X'_{2t})$ and $Y_t=(Y'_{1t}, Y'_{2t})'$ and let us assume that the processes (X'_{1t}, Y'_{1t}) and (X'_{2t}, Y'_{2t}) are independent; then we have:

$$C_{X\to Y}=C_{X1\to Y1}+C_{X2\to Y2}$$

and similar results for the other measures.

Proof: Independence implies that the dominating measure μ_y can be supposed to be a product measure $\mu_y^1 \otimes \mu_y^2$ (possibly depending on t) and the conditional p. d. f. are such that:

$$l_{0t}(y_t/x_{t-1}^{t-p}, y_{t-1}^{t-p})=l_{0t}(y_{1t}/x_{1,t-1}^{t-p}, y_{1,t-1}^{t-p})l_{0t}(y_{2t}/x_{2,t-1}^{t-p}, y_{2,t-1}^{t-p}).$$

and

$$l_{0t}(y_t/y_{t-1}^{t-p})=l_{0t}(y_{1t}/y_{1,t-1}^{t-p})\times l_{0t}(y_{2t}/y_{2,t-1}^{t-p})$$

The equality of the theorem follows immediately. $\square$

5 Estimation of the Causality Measures and Hypotheses Testing

5.a. The Observations

Two cases are of interest. In the first case we have only one observation of the variables Z_t, $t=-p+1, \ldots, T$. In order to be able to make inference we assume that the process is stationary, or at least homogeneous. We know that, in this case, the discrepancies and causality measures have special forms; moreover asymptotic arguments can be used as T goes to infinity.

Another interesting case is that in which we have I independent observations of the process Z; i. e. at each date t we observe I independent variables $Z_t^i\, i=1, \ldots, I$. In such a case it is not necessary to assume stationarity or homogeneity and the asymptotic arguments assume that I goes to infinity while T is fixed.

In any case the p. d. f. of the observations can be written:

$$\prod_{i=1}^{I} l_{0T}(\mathbf{z}_T^i),$$

where $\mathbf{z}_T^i$ is the i-th observation of $\mathbf{Z}_T=(Z'_T, \ldots, Z'_{-p+1})'$.

5.b. Estimation of the Discrepancies

Let us now assume that the p. d. f. of the observations depends on a vector of parameters θ; this p. d. f. is denoted by: $\prod_{i=1}^{I} l_T(\mathbf{z}_T^i, \theta)$. It is also possible to define the log-likelihood function:

$$L_N(\theta)=\sum_{i=1}^{I} \operatorname{Log} l_T(\mathbf{z}_T^i, \theta),$$

where $N=TI$. Note that in both cases considered in 5.*a*, the asymptotic properties will be studied when N goes to infinity.

Let us consider an hypothesis $\tilde{H}$ defined by $h(\theta)=0$. According to subsection 4.*a*, the discrepancy between the true p. d. f. $l_T(\mathbf{z}_T, \theta_0)$ and H is:

$$\tilde{D}=\frac{1}{T}\operatorname*{Min}_{\theta\in\tilde{H}} E_0 \operatorname{Log}\frac{l_T(\mathbf{Z}_T, \theta_0)}{l_T(\mathbf{Z}_T, \theta)}$$

or

$$\tilde{D}=\frac{1}{T}E_0 \operatorname{Log}\frac{l_T(\mathbf{Z}_T, \theta_0)}{l_T[\mathbf{Z}_T, \tilde{\theta}(\theta_0)]}$$

where $\tilde{\theta}(\theta_0)$ is the pseudo-true value of θ.

Let us denote by $\hat{\theta}$ and $\hat{\tilde{\theta}}$ the unconstrained and the constrained maximum likelihood estimators of θ.

A natural estimator of D is:

$$\hat{\tilde{D}}^1=\frac{1}{N}[L_N(\hat{\theta})-L_N[\tilde{\theta}(\hat{\theta})]].$$

It is clear that in the contexts considered in 5.*a*, and for the discrepancies considered in 4.*b*, this estimator is consistent. However another estimator, based on the likelihood ratio statistic LR, could be proposed:

$$\hat{\tilde{D}}^2 = \frac{1}{N}[L_N(\hat{\theta}) - L_N(\hat{\tilde{\theta}})] = \frac{1}{2N} LR.$$

This estimator is also consistent and the following theorem shows how $\hat{\tilde{D}}^1$ and $\hat{\tilde{D}}^2$ are linked under $\tilde{H}$.

THEOREM 8 : Under $\tilde{H}$, we have

$$\hat{\tilde{D}}^1 - \hat{\tilde{D}}^2 = o_p(N^{-1}).$$

Proof: See appendix 4.

The previous theorem has an important corollary:

COROLLARY 9 :

$$2N\hat{\tilde{D}}^1 = 2[L_N(\hat{\theta}) - L_N[\tilde{\theta}(\hat{\theta})]]$$

is asymptotically equivalent, under $\tilde{H}$, to the likelihood ratio statistic $2N\hat{\tilde{D}}^2$.

Proof: Immediate.

5.*c*. Estimation of Causality Measures and Hypotheses Testing

We still assume that the true p. d. f. belongs to a family indexed by θ. Moreover we also assume that all the pseudo p. d. f. associated with the various causality hypotheses also, belong to this family. The basic non causality hypotheses are now defined by:

H_1: non causality from X to Y: $h_1(\theta) = 0$;

H_2: non causality from Y to X: $h_2(\theta) = 0$;

H_3: non instantaneous causality: $h_3(\theta) = 0$.

The seven discrepancies D_1, D_2, D_3, D_{12}, D_{13}, D_{23}, D_{123} now depend on θ, and they can be consistently estimated by using the results of the previous subsection.

Let us for instance consider $D_1(\theta)$; it can be consistently estimated by

$$\hat{D}_1^1 = \frac{1}{N}[L_N(\hat{\theta}) - L_N[\tilde{\theta}(\hat{\theta})]]$$

$$\hat{D}_1^1 = \frac{1}{TI}\sum_{i=1}^{I}\sum_{t=1}^{T} \text{Log}\frac{l_t^i(y_t/x_{t-1}^{t-p}, y_{t-1}^{t-p}; \hat{\theta})}{l_t^i(y_t/y_{t-1}^{t-p}; \hat{\theta})}$$

Similarly we obtain $\hat{D}_1^1$, $\hat{D}_3^1$, $\hat{D}_{12}^1$, $\hat{D}_{23}^1$, $\hat{D}_{123}^1$.

From the proofs of theorems 2 and 3, it is clear that the estimators of the causality measures are:

$$\hat{C}^1_{X \to Y} = \hat{D}_1^1 = \hat{D}_{13}^1 - \hat{D}_3^1 = \hat{D}_{123}^1 - \hat{D}_{23}^1$$
$$\hat{C}^1_{Y \to X} = \hat{D}_2^1 = \hat{D}_{23}^1 - \hat{D}_3^1 = \hat{D}_{123}^1 - \hat{D}_{13}^1$$

and

$$\hat{C}^1_{X \quad Y} = \hat{D}_3^1 = \hat{D}_{13}^1 - \hat{D}_1^1 = \hat{D}_{23}^1 - \hat{D}_2^1.$$

These equalities and theorem 8 have important consequences for the procedures of hypotheses testing. Let us consider for instance the equality.

$$\hat{D}_1^1 = \hat{D}_{13}^1 - \hat{D}_3^1.$$

From theorem 8 we have, under H_{13}:

$$\hat{D}_1^2 = \hat{D}_{13}^2 - \hat{D}_3^2 + o_p(N^{-1})$$
$$LR_1 = LR_{13} - LR_3 + o_p(1)$$

where LR_1 is the likelihood ratio statistic for testing H_1 against the maintained hypothesis H and LR_{13}, LR_3 have similar meanings.

The statistic $LR_{13} - LR_3$ is the likelihood ratio statistic for testing H_{13} against H_3. Therefore, the previous equality shows that the likelihood ratio test statistics for testing H_1 against H or H_{13} against H_3 are asymptotically equivalent under H_{13}.

Moreover, if we consider the sequence of nested hypotheses $H \supset H_3 \supset H_{13}$, it is well known that, under H_{13}, the LR statistic for testing H_3 against H is asymptotically independent from the LR statistic for testing H_{13} against H_3 in other words LR_3 and $LR_{13} - LR_3$ are asymptotically independent under H_{13}. Since $LR_{13} - LR_3 = LR_1 + o_p(1)$, we see that LR_1 and LR_3 are also asymptotically independent under H_{13}.

Obviously there are many other results of this kind which can be summarized in the following theorem.

THEOREM 10: Let (i, j, k) be one of the four ordered sets (1, 3, 2), (2, 3, 1), (3, 1, 2), (3, 2, 1).

The LR statistic for testing H_j against H and H_{ij} against H_i are asymptotically equivalent under H_{ij} and the LR statistics LR_i and LR_j (for testing respectively H_i and H_j against H) are asymptotically independent under H_{ij}.

The LR statistic for testing H_k against H and H_{ijk} against H_{ij} are asymptotically equivalent under H_{ijk} and the LR statistic LR_k and LR_{ij} are asymptotically independent under H_{ijk}. Moreover, for the first two sets, the LR statistic for testing H_{jk} against H and H_{ijk} against H_i are asymptotically equivalent under H_{ijk} and the LR statistic LR_i and LR_{jk} are asymptotically independent under H_{ijk}.

Another way of looking at this result is the following. Let us consider two hypotheses $H_a H_b$ and their intersection $H_c = H_a \cap H_b$, and let us assume that $D_c = D_a + D_b$ (note that all the equalities of definitions 1, 2 and 3 are of this form); the LR statistic for testing H_a against H is equivalent, under H_c, to the LR statistic for testing H_c against H_b; moreover the LR statistics for testing respectively H_a against H and H_b against H are *asymptotically independent* under H_c. This last property implies that for testing H_c against H is possible to test H_a against H, at a given level α_a, to test H_b against H at a leval α_b and to accept H_c if both H_a and H_b are accepted; the overall level α of the test is given by $1-\alpha=(1-\alpha_a)(1-\alpha_b)$ or $\alpha=\alpha_a+\alpha_b-\alpha_a\alpha_b$.

The basic equality

$$C_{X,Y}=C_{X\to Y}+C_{Y\to X}+C_{X\leftrightarrow Y}$$

or

$$D_{123}=D_1+D_2+D_3$$

also shows, in the same way, that the test of independence can be split into three independent tests based on the statistics $2N\hat{C}_{X\to Y}$, $2N\hat{C}_{Y\to X}$ and $2N\hat{C}_{X\leftrightarrow Y}$ which are asymptotically independent under H_{123}. These statistics and their sum $\hat{C}_{X,Y}$ are asymptotically equivalent to the corresponding likelihood ratio statistics.

Similarly $2N\hat{C}_{X\to Y}$ and $2N\hat{C}_{X\leftrightarrow Y}$ are asymptotically independent under H_{13}; symmetrically $2N\hat{C}_{Y\to X}$ and $2N\hat{C}_{X\leftrightarrow Y}$ are asymptotically independent under H_{23} However $2N\hat{C}_{X\to Y}$ and $2N\hat{C}_{Y\to X}$, though they are asymptotically independent under H_{123}, are not, in general, asymptotically independent under $H_{12}-H_{123}$.

As far as Sims' causality measures and test are concerned similar remarks can be made. If panel data are available, the definitions of Sims' causality measure and Markov measure are given by definitions 5 and 6.

$CS_{X\to Y}$ can be estimated by

$$\hat{C}S^1_{X\to Y}=\frac{1}{N}[L_N(\hat{\theta})-L_N[\tilde{\theta}(\hat{\theta})]$$

$$=\frac{1}{TI}\sum_{t=1}^{T}\sum_{i=1}^{I}\text{Log}\frac{l_t^i(x_t/x_{t-1}^{t-p},\ y_T^{t-p};\ \hat{\theta})}{l_t^i(x_t/x_{t-1}^{t-p},\ y_t^{t-p},\ \hat{\theta})}.$$

or

$$\hat{C}S^2_{X\to Y}=\frac{1}{N}[L_N(\hat{\theta})-L_N(\hat{\tilde{\theta}}_n)]$$

where $\hat{\tilde{\theta}}_n$ is the estimator constrained by the non causality from X to Y in Sims' sense.

Similarly M_Y can be estimated by:

$$\hat{M}_Y^1=\frac{1}{TI}\sum_{i=1}^{T}\sum_{t=1}^{I}\text{Log}\frac{l_t^i(y_t/x_0^{1-p},\ y_{t-1}^{1-p};\ \hat{\theta})}{l_t^i(y_t/y_{t-1}^{t-p},\ \hat{\theta})}$$

or

$$\hat{M}_Y^2 = \frac{1}{N}[L_N(\hat{\theta}) - L_N(\hat{\hat{\theta}}_N)]$$

where $\hat{\hat{\theta}}_N$ is the ML estimator constrained by the nullity of the markovian measure.

The test statistics are respectively, $2N\hat{CS}^1_{X\to Y}$ or $2N\hat{CS}^2_{X\to Y}$ and $\hat{M}^1_Y$ or $\hat{M}^2_Y$.

Theorem 4 shows, for instance, that the test of no causality in Granger sense can be split up into two asymptotically independent tests: a test of non causality in Sims' sense based on $2N\hat{CS}^1_{X\to Y}$ (or $2\hat{N}CS^2_{X\to Y}$) and a test of markovian property of Y based on $2N\hat{M}^1_Y$ or $2N\hat{M}^2_Y$.

If only one path of the process has been observed and if homogeneity is assumed, the definitions $CS^*_{X\to Y}$ and M^*_Y of theorem 5 have to be used. However since infinite set of variables appear in these definitions, truncations have to be made (see the autoregressive exemple below).

6 Applications

6.*a*. Vector Autoregressive Models

i) ***Causality measures*** The previous results may be applied, in particular, when the process $Z = (X', Y')'$ is a gaussian process with an autoregressive representation of order p. In such a case the various conditional distributions appearing in the causality measures are normal and the various pseudo true p.d.f. belong to the initial familly. Moreover, if Z has an autoregressive representation with roots outside of the unit cricle, this process is regular stationary and assumption (A) is satisfied. Let us also assume that only one path of the process is observed ($I = 1$ and $N = T$).

Note that for a Gaussian p.d.f.:

$$l_0(u) = \frac{1}{(2\pi)^{J/2}} \frac{1}{\sqrt{\det \Sigma}} \exp - \frac{1}{2}(u-m)' \Sigma^{-1} (u-m),$$

we have :

$$E_0 \operatorname{Log} l_0(U) = -\frac{J}{2}\operatorname{Log} 2\pi - \frac{1}{2}\operatorname{Log}\det\Sigma - \frac{J}{2}$$

and, consequently, all the causality measures can be expressed in terms of logarithms of covariance matrix determinants. To obtain the exact expressions of the discrepancies, we introduce the following regressions using Geweke's notations [GEWEKE [1982]]:

Initial A.R. model

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \sum_{s=1}^{p} B_s \begin{bmatrix} X_{t-s} \\ Y_{t-s} \end{bmatrix} + \varepsilon_t, \qquad E\,\varepsilon_t = 0, \qquad V\,\varepsilon_t = \gamma$$

Finite marginal autoregressions

$$X_t = \sum_{s=1}^{p} E_{1s} X_{t-s} + u_{1t}, \qquad E\,u_{1t} = 0, \qquad V\,u_{1t} = \Sigma_1,$$

$$Y_t = \sum_{s=1}^{p} G_{1s} Y_{t-s} + v_{1t}, \qquad E\,v_{1t} = 0, \qquad V\,v_{1t} = T_1.$$

Infinite marginal autoregressions

$$X_t = \sum_{s=1}^{\infty} \bar{E}_{1s} X_{t-s} + \bar{u}_{1t}, \qquad E\,\bar{u}_{1t} = 0, \qquad V\,\bar{u}_{1t} = \bar{\Sigma}_1,$$

$$Y_t = \sum_{s=1}^{\infty} \bar{G}_{1s} Y_{t-s} + \bar{v}_{1t}, \qquad \bar{E}\,v_{1t} = 0, \qquad V\,\bar{v}_{1t} = \bar{T}_1$$

Finite joint autoregressions

$$X_t = \sum_{s=1}^{p} E_{2s} X_{t-s} + \sum_{s=1}^{p} F_{2s} Y_{t-s} + u_{2t}, \qquad E\,u_{2t} = 0, \qquad V\,u_{2t} = \Sigma_2,$$

$$Y_t = \sum_{s=1}^{p} G_{2s} Y_{t-s} + \sum_{s=1}^{p} H_{2s} X_{t-s} + v_{2t}, \qquad E\,v_{2t} = 0, \qquad V\,v_{2t} = T_2.$$

Finite joint autogressions including present values

$$X_t = \sum_{s=1}^{p} E_{3s} X_{t-s} + \sum_{s=0}^{p} F_{3s} Y_{t-s} + u_{3t}, \qquad E\,u_{3t} = 0, \qquad V\,u_{3t} = \Sigma_3,$$

$$Y_t = \sum_{s=1}^{p} G_{3s} Y_{t-s} + \sum_{s=0}^{p} H_{3s} X_{t-s} + v_{3t}, \qquad E\,v_{3t} = 0, \qquad V\,v_{3t} = T_3,$$

Theorem 5 shows that the causality measures in Sims' sense will be based on the following infinite regressions.

Autoregressions "à la Sims"

$$X_t = \sum_{s=1}^{p} E_{4s} X_{t-s} + \sum_{s=-\infty}^{p} F_{4s} Y_{t-s} + u_{4t}, \qquad E\,u_{4t}=0, \qquad V\,u_{4t}=\Sigma_4,$$

$$Y_t = \sum_{s=1}^{p} G_{4s} Y_{t-s} + \sum_{s=-\infty}^{p} H_{4s} X_{t-s} + v_{4t}, \qquad E\,v_{4t}=0, \qquad V\,v_{4t}=T_4,$$

The causality measures are functions of the variance-covariance matrices γ, Σ, T of the error terms. They are given by:

$$C_{X \to Y} = \frac{1}{2} \mathrm{Log} \frac{\det T_1}{\det T_2}, \qquad C_{Y \to X} = \frac{1}{2} \mathrm{Log} \frac{\det \Sigma_1}{\det \Sigma_2},$$

$$C_{X \leftrightarrow Y} = \frac{1}{2} \mathrm{Log} \frac{\det \Sigma_2 \det T_2}{\det \gamma}$$

$$= \frac{1}{2} \mathrm{Log} \frac{\det \Sigma_2}{\det \Sigma_3} = \frac{1}{2} \mathrm{Log} \frac{\det T_2}{\det T_3},$$

$$C_{X,Y} = \frac{1}{2} \mathrm{Log} \frac{\det \Sigma_1 \det T_1}{\det \gamma}$$

$$CS^*_{X \to Y} = \frac{1}{2} \mathrm{Log} \frac{\det \Sigma_3}{\det \Sigma_4}, \qquad CS^*_{Y \to X} = \frac{1}{2} \mathrm{Log} \frac{\det T_3}{\det T_4},$$

$$CS^*_{X \Rightarrow Y} = \frac{1}{2} \mathrm{Log} \frac{\det \Sigma_2}{\det \Sigma_4}, \qquad CS^*_{Y \Rightarrow X} = \frac{1}{2} \mathrm{Log} \frac{\det T_2}{\det T_4},$$

$$M^*_Y = \frac{1}{2} \mathrm{Log} \frac{\det T_1}{\det \bar{T}_1}, \qquad M^*_X = \frac{1}{2} \mathrm{Log} \frac{\det \Sigma_1}{\det \bar{\Sigma}_1}.$$

These measures correspond to the measures proposed by Geweke, with a multiplicative factor $\frac{1}{2}$.

It is interesting to note that the assumption of an autoregressive representation of order p, i. e. the markovian assumption, has been explicitly taken into account. In particular the results of the previous sections explain how the various regressions have to be truncated. For example, despite the fact that process Y does not have a marginal autoregressive representation of order p, the marginal regression has to be performed after a truncation at this order p. This may be viewed as a justification of the procedure proposed in Geweke, where the theoretical results are derived without assuming a Markov property for the process and where all the regressions are truncated at the same order p.

However it is clear that this practice is incorrect in the case of regressions "à la Sims", where the truncation orders for the past and the future are not symmetric [p for the past, ∞ for the future]. Moreover theorem 4 gives the relation between Granger's causality measures and causality measures "à la Sims", for instance:

$$\begin{aligned} C_{X \to Y} &= CS^*_{X \to Y} + M^*_Y \\ \Leftrightarrow \mathrm{Log} \frac{\det T_1}{\det T_2} &= \mathrm{Log} \frac{\det \Sigma_3}{\det \Sigma_4} + \mathrm{Log} \frac{\det T_1}{\det \bar{T}_1}. \end{aligned}$$

This relation generalises the one derived in GEWEKE [1982] Theorem 1. In effect, the case of a stationary regular process with an infinite autoregressive representation is obtained as a limit case when p tends to infinity. In this context the measure M^*_Y tends to zero and the relation is exactly Geweke's:

$$\mathrm{Log} \frac{\det T_1}{\det T_2} = \mathrm{Log} \frac{\det \Sigma_3}{\det \Sigma_4}.$$

However this relation has to be modified when the finite character of the autoregression is explicitly taken into account. Finally note that the proof of this relation given in appendix 3 is in time domain and differs from Geweke's proof, which is in frequency domain.

ii) *Estimation of the Causality Measures and Hypotheses Testing*

We deduce from section 5 and in particular from theorem 8, that estimators of $C_{X \to Y}$, $C_{Y \to X}$, $C_{X \leftrightarrow Y}$ are easily obtained by replacing in these causality measures the theoretical covariance matrices by their empirical counterpart based on the ordinary least squares residuals. If we denote by K and L the respective sizes of the processes X and Y by $\hat{\gamma}$, $\hat{\Sigma}$, $\hat{T}$ the empirical covariance matrices, the likelihood ratio test statistics and the associated degrees of freedom of their chi-square distribution under the null hypothesis are given below

TABLE 3

Hypothesis to be tested	Test statistics	Degrees of freedom
H_1 (X does not cause Y)	$T\,\mathrm{Log} \frac{\det \hat{T}_1}{\det \hat{T}_2}$	KLp
H_2 (Y does not cause X)	$T\,\mathrm{Log} \frac{\det \hat{\Sigma}_1}{\det \hat{\Sigma}_2}$	KLp
H_3 (no instantaneous causality)	$T\,\mathrm{Log} \frac{\det \hat{\Sigma}_2}{\det \hat{\Sigma}_3}$ or $T\,\mathrm{Log} \frac{\det \hat{T}_2}{\det \hat{T}_3}$ or $T\,\mathrm{Log} \frac{\det \hat{\Sigma}_2 \det \hat{T}_2}{\det \hat{\gamma}}$	KL
H_{123} (independence)	$T\,\mathrm{Log} \frac{\det \hat{\Sigma}_1 \det \hat{T}_1}{\det \hat{\gamma}}$	$KL(2p+1)$

The general results of section 5 apply. In particular the three statistics for testing H_1, H_2 and H_3 against H are asymptotically independent under H_{123}. Note however that the first two statistics are not independent under $H_{12}-H_{123}$.

Finally we can remark that tests of the Markov property or of Sims' non causality are difficult to perform in this case. In effect for these hypotheses some of the regressions contain an infinite number of terms. It is of course possible to use truncations, however the degrees of freedom of the associated chi-square distribution and the results of the tests procedures depend on the choice of these truncation levels.

6.*b.* Qualitative Processes: Observation of a Stationary Markov Chain

i) *Causality Measures*

Another application concerns qualitative processes. Let us consider an homogeneous Markov chain of order one. The number of possible states is finite and the general element of the transition matrix is:

$$P[X_t=i,\ Y_t=k/X_{t-1}=j,\ Y_{t-1}=l]=p(i,\ k/j,\ l),$$
$$i,\ j=1\ldots J,\qquad k,\ l=1\ldots L,$$

and the element of the invariant initial measure is:

$$P[X_t=i,\ Y_t=k]=\pi(i,\ k),\qquad i=1\ldots J,\quad k=1\ldots L.$$

We deduce the expressions of the conditional densities appearing in the causality measures; we have:

$$P[X_t=i/X_{t-1}=j,\ Y_{t-1}=l]=\sum_{k=1}^{L} p(i,\ k/j,\ l)=p(i,.\,/j,\ l),$$

$$P[Y_t=k/X_{t-1}=j,\ Y_{t-1}=l]=\sum_{i=1}^{J} p(i,\ k/j,\ l)=p(.,\ k/j,\ l),$$

$$P[X_t=i/X_{t-1}=j)=\frac{P(X_t=i,\ X_{t-1}=j)}{P(X_{t-1}=j)}$$

$$=\frac{\sum_{k=1}^{L}\sum_{l=1}^{L} p(i,\ k/j,\ l)\,\pi(j,\ l)}{\sum_{l=1}^{L}\pi(j,\ l)}$$

$$=p_x(i/j)\quad\text{[say]}.$$

Similarly we have:

$$P[Y_t=k/Y_{t-1}=l]=\frac{\sum_{i=1}^{J}\sum_{j=1}^{J}p(i,k/j,l)\pi(j,l)}{\sum_{j=1}^{J}\pi(j,l)}=p_y(k/l)\quad [\text{say}]$$

The various measures are now given by:

$$C_{X\to Y}=\sum_{j=1}^{J}\sum_{l=1}^{L}\pi(j,l)\left\{\sum_{k=1}^{L}p(.,k/j,l)\operatorname{Log}\frac{p(.,k/j,l)}{p_y(k/l)}\right\},$$

$$C_{Y\to X}=\sum_{j=1}^{J}\sum_{l=1}^{L}\pi(j,l)\left\{\sum_{i=1}^{J}p(i,./j,l)\operatorname{Log}\frac{p(i,./j,l)}{p_x(i/j)}\right\},$$

$$C_{X\leftrightarrow Y}=\sum_{j=1}^{J}\sum_{l=1}^{L}\pi(j,l)\left\{\sum_{i=1}^{J}\sum_{k=1}^{L}p(i,k/j,l)\operatorname{Log}\frac{p(i,k/j,l)}{p(i,./j,l)p(.,k/j,l)}\right\}$$

$$C_{X,Y}=\sum_{j=1}^{J}\sum_{l=1}^{L}\pi(j,l)\left\{\sum_{i=1}^{J}\sum_{k=1}^{L}p(i,k/j,l)\operatorname{Log}\frac{p(i,k/j,l)}{p_x(i/j)p_y(k/l)}\right\}.$$

These measures appear as convex combinations of disagregate causality measures computed conditionally on the previous state:

$$C_{X\to Y}=\sum_{j=1}^{J}\sum_{l=1}^{L}\pi(j,l)C_{X\to Y}(j,l),$$

$$C_{Y\to X}=\sum_{j=1}^{J}\sum_{l=1}^{L}\pi(j,l)C_{Y\to X}(j,l),$$

$$C_{X\leftrightarrow Y}=\sum_{j=1}^{J}\sum_{l=1}^{L}\pi(j,l)C_{X\leftrightarrow Y}(j,l),$$

$$C_{X,Y}=\sum_{j=1}^{J}\sum_{l=1}^{L}\pi(j,l)C_{X,Y}(j,l).$$

It is easily seen that:

$$\forall j,l:\quad C_{X,Y}(j,l)=C_{X\to Y}(j,l)+C_{Y\to X}(j,l)+C_{X\leftrightarrow Y}(j,l).$$

These disaggregate measures are useful for the study of the transmission of the shocks through the various states. They might lead to a classification of the states in terms of causality.

ii) ***Chi-Square Causality Measures***

In the case of qualitative models, Kullback information measures are often replaced by empirical chi-squares. Among the possible chi-square approximations, only some particular choices are compatible with a decomposition of the from:

$$\chi^2_{X,\,Y}(j,\,l)=\chi^2_{X\leftrightarrow Y}(j,\,l)+\chi^2_{X\rightarrow Y}(j,\,l)+\chi^2_{Y\rightarrow X}(j,\,l).$$

For instance we have the following property:

THEOREM 11 : The chi-square measures:

$$\chi^2_{X,\,Y}(j,\,l)=\sum_{i=1}^{J}\sum_{k=1}^{L}\frac{[p(i,\,k/j,\,l)-p_x(i/j)\,p_y(k/l)]^2}{p_x(i/j)\,p(.,\,k/j,\,l)},$$

$$\chi^2_{X\leftrightarrow Y}(j,\,l)=\sum_{i=1}^{J}\sum_{k=1}^{L}\frac{[p(i,\,k/j,\,l)-p(i,\,./j,\,l)\,p(.,\,k/j,\,l)]^2}{p_x(i/j)\,p(.,\,k/j,\,l)},$$

$$X^2_{X\rightarrow Y}(j,\,l)=\sum_{k=1}^{L}\frac{[p_y(k/l)-p(.,\,k/j,\,l)]^2}{p(.,\,k/j,\,l)},$$

$$X^2_{Y\rightarrow X}(j,\,l)=\sum_{i=1}^{J}\frac{[p_x(i/j)-p(i,\,./j,\,l)]^2}{p_x(i/j)},$$

satisfy the decomposition above.
Proof: See appendix 5.

The proof of theorem 11 shows that there exist some other possible choices of chi-square causality measures satisfying the decomposition. For instance another choice is obtained by symmetry, replacing the denominator $p_x(i/j)\,p(.,\,k/j,\,l)$ by $p(i,\,./j,\,l)\,p_y(k/l)$. The first choice appears interesting if X is thought to be exogenous, the second one when Y is thought to be exogenous.

It is also easily verified that a similar decomposition does not exist when the denominator is replaced by $p_x(i/j)\,p_y(k/l)$ or by $p(i,\,./j,\,l)\,p(.,\,k/j,\,l)$.

iii) ***Hypotheses Testing***

Let us assume that the process (X, Y) is observed between 1 and T and let us denote:

$$p(i,\,k/j,\,1)=\frac{n(i,\,k;\,j,\,l)}{n(j,\,l)}$$

the ratio between the number of transitions from state $(j,\,l)$ to state $(i,\,k)$ and the number of dates at which $X_t=j$ and $Y_t=l$, $\hat{\pi}(i,\,k)=\dfrac{n(j,\,k)}{T}$ the empirical frequency of state $(i,\,k)$.

Test statistics are deduced from the causality measures $C_{X\rightarrow Y}$, $C_{Y\rightarrow X}$, $C_{X\leftrightarrow Y}$ after replacement of the probabilities $p(i,\,k/j,\,l)$; $\pi(i,\,k)$ by their empirical counterparts $\hat{p}(i, k/j, l)$ and $\hat{\pi}(i, k)$. For instance, for testing non causality from X to Y, we can use:

$$2\,\mathrm{T}\hat{\mathrm{C}}^{1}_{\mathrm{X}\to\mathrm{Y}} = 2\,\mathrm{T}\sum_{j=1}^{\mathrm{J}}\sum_{l=1}^{\mathrm{L}}\hat{\pi}(j, l)\left\{\sum_{k=1}^{\mathrm{L}}\hat{p}(., k/j, l)\,\mathrm{Log}\frac{\hat{p}(., k/j, l)}{\hat{p}_y(k/l)}\right\},$$

where: $\hat{p}(., k/j, l) = \sum_{i=1}^{\mathrm{J}} \hat{p}(i, k/j, l)$

$$\hat{p}_y(k/l) = \frac{\sum_{i=1}^{1}\sum_{j=1}^{\mathrm{J}}\hat{p}(i, k/j, l)\,\hat{\pi}(j, l)}{\sum_{j=1}^{\mathrm{J}}\hat{\pi}(j, l)}.$$

It is of course possible to replace this statistic by its chi-square approximation, which is equivalent under the null.

Finally we give below the degrees of freedom associated with the different causality hypotheses.

Hypothesis	Degrees of freedom
X does not cause Y.	L(J−1)(L−1)
Y does not cause X.	J(J−1)(L−1)
Absence of instantaneous causality.	JL(J−1)(L−1)
Independence.	JL(JL−1)−J(J−1)−L(L−1)

The degrees of freedom are such that

$$(\mathrm{J}-1)(\mathrm{L}-1))(\mathrm{JL}+\mathrm{L}+\mathrm{J})) = \mathrm{JL}(\mathrm{JL}-1)-[\mathrm{J}(\mathrm{J}-1)+\mathrm{L}(\mathrm{L}-1)].$$

6.*c* Qualitative Processes: Observations of Non Homogenous Markov Chains
(compare Bouissou-Laffont-Vuong [1986])

If we relax the homogeneity assumption, the transition probabilities are now time dependent and denoted by:

$$\mathrm{P}[\mathrm{X}_t = i,\ \mathrm{Y}_t = k/\mathrm{X}_{t-1} = j,\ \mathrm{Y}_{t-1} = 1] = p_t(i, k/j, l).$$

we deduce:

$$\mathrm{P}[\mathrm{X}_t = i/\mathrm{X}_{t-1} = j,\ \mathrm{Y}_{t-1} = j,\ \mathrm{Y}_{t-1} = l] = \sum_k p_t(l, k/j, l) = p_t(i, ./j, l)$$

and

$$P[Y_t = k/X_{t-1} = j,\ Y_{t-1} = l] = p_t(.,\ k/j,\ l).$$

The marginal distribution of (X_t, Y_t) depends on the initial marginal distribution π_0 and on the transition probabilities associated with the dates smaller than t. If this marginal distribution is denoted by:

$$P[X_t = i,\ Y_t = k] = \pi_t(i, k),$$

we have for instance for the Kullback measure of instantaneous causality:

$$C_{X \leftrightarrow Y} = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J} \sum_{l=1}^{L} \pi_{t-1}(j, l)$$

$$\times \left\{ \sum_{i=1}^{J} \sum_{k=1}^{L} p_t(i, k/j, l) \operatorname{Log} \frac{p_t(i, k/j, l)}{p_t(i, ./j, l)\, p_t(., k/j, l)} \right\}$$

$$= \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J} \sum_{l=1}^{L} \pi_{t-1}(j, l)\, C_{X \leftrightarrow Y}(t; j, l),$$

where $C_{X \leftrightarrow Y}(t; j, l)$ is a causality measure associated with state j, l and transition between $t-1$ and t. These time disagregated measures $C_{X \leftrightarrow Y}(t; j, l)$ are useful to examine the evolution of causality during the observation period.

The measures can be easily estimated from panel data, by replacing the theoretical probabilities by their empirical counterparts.

If I independent realisations of the chain are observable, the statistics:

$$2 I \hat{\pi}_t(j, l)\, \hat{C}_{X \leftrightarrow Y}(t; j, l) = 2 I\, \hat{\pi}_t(j, l) \sum_{i=1}^{J} \sum_{k=1}^{L} \hat{p}_t(i, k/j, l)$$

$$\times \operatorname{Log} \frac{\hat{p}_t(i, k/j, l)}{\hat{p}_t(i, ./j, l)\, \hat{p}_t(., k/j, l)}$$

$$= 2 I\, \pi_t(j, l)\, \hat{C}_{X \leftrightarrow Y}(t; j, l) + o_p(1)$$

have asymptoticaly $(I \to \infty)$, under the null hypothesis of non instantaneous causality, a chi-square distribution with $(J-1)(L-1)$ degrees of freedom, and they are asymptotically independent [see Anderson-Goodman [1957] and Gourieroux [1984] Chapter V].

The global test for non instantaneous causality should be based on :

$$2 T I \hat{C}_{X \leftrightarrow Y} = \sum_{t=1}^{T} \sum_{j=1}^{J} \sum_{l=1}^{L} 2 I\, \hat{\pi}_{t-1}(j, l)\, \hat{C}_{X \leftrightarrow Y}(t; j, l)$$

Under the null hypothesis this statistic has an asymptotic chi-square distribution with $TJL(J-1)(L-1)$ degrees of freedom.

Similar decompositions may also be derived for the other causality measures. For instance we have:

$$C_{Y \to X} = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J} \sum_{l=1}^{L} \pi_{t-1}(j, l) \left[\sum_{i=1}^{J} p_t(i, . /j, l) \operatorname{Log} \frac{p_t(i, . /j, l)}{\hat{p}_{x,t}(i/j)} \right]$$

$$= \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J} \sum_{l=1}^{L} \pi_{t-1}(j, l)\, C_{Y \to X}(t; j, l).$$

The associated test statistic obtained by replacing the theoretical probabilities by the empirical ones and by multiplying by 2 TI has asymptotically a chi-square distribution with $TJ(J-1)(L-1)$ degrees of freedom.

7 Concluding Remarks

In this paper we have proposed a general framework for treating the causality relations between two sets of variables. The case of more than two sets of variables is beyond the scope of the present paper and is left to further research; in particular if there are more than two sets of variables it is probably necessary to distinguish between direct and indirect kinds of causality; it is an open question to know if these problems can be tackled with the methods presented here.

APPENDIX 0

Markovian Property of the Marginal Process

In order to compute the discrepancies between $l_{0\,T}$ and any of the hypotheses $\tilde{H}$, which is a subset of H_1 or H_2, it is useful to prove first the following lemma:

LEMMA : Under the hypothesis of non causality from X to Y H_1 (resp. H_2) the process Y_t (resp X_t) is markovian of order p.

Proof:

$$l_{0\,t}(y_t/y_{t-1}^{t-p}) = \int_{\mathscr{X}^p} l_{0\,t}(y_t/x_{t-1}^{t-p}, y_{t-1}^{t-p})\, l_{0\,t}(x_{t-1}^{t-p}/y_{t-1}^{t-p}) \bigotimes_{i=t-p}^{t-1} \mu_x(dx_i)$$

But, under H_1:

$$l_{0\,t}(y_t/x_{t-1}^{t-p}, y_{t-1}^{t-p}) = l_{0\,t}(y_t/\mathbf{x}_{t-1}, \mathbf{y}_{t-1}) = l_{0\,t}(y_t/\mathbf{y}_{t-1})$$

Hence:

$$l_{0\,t}(y_t/y_{t-1}^{t-p}) = l_{0\,t}(y_t/\mathbf{y}_{t-1}) \int_{\mathscr{X}^p} l_{0\,t}(x_{t-1}^{t-p}/y_{t-1}^{t-p}) \bigotimes_{i=t-p}^{t-1} \mu_x(dx_i)$$

$$= l_{0\,t}(y_t/\mathbf{y}_{t-1})$$

which proves that Y_t is markovian of order p. □

It is worth noting that this proof crucially uses the assumption that the p. d. f. $l_{0\,t}(./.)$ are sefined with respect to a measure μ which is a product $\mu_x \otimes \mu_y$.

If it is not the case, the result of the lemma may be incorrect as in the following example:

$$\begin{cases} Y_t = Y_{t-1} - 1/4\, Y_{t-2} + u_t \\ X_t = Y_{t-1} \end{cases}$$

where u_t is a scalar white noise.

It is clear that $Z_t = (X_t, Y_t)'$ is markovian of order 1 and that X does not cause Y, but Y_t is only markovian of order 2 (and not of order 1).

APPENDIX 1

Derivations of the Causality Measures

With obvious consideration of symmetry, it is easy to deduce the form of D_{23} (resp. D_2) from that of D_{13} (resp. D_1). Thus, we only derive in this appendix the three discrepancies D_1, D_3, D_{123}.

a) *Derivation of* D_1

Under H_1, the p. d. f.; of $\mathbf{Z}_T$ has the following form :

$$l(\mathbf{z}_T) = f(\mathbf{x}_0, \mathbf{y}_0) \prod_{t=1}^{T} l_t(x_t/\mathbf{x}_{t-1}, \mathbf{y}_t)\, l_t(y_t/\mathbf{y}_{t-1}).$$

Since, under H_1, the process Y_t is markovian of order p, it follows that:

$$D_1 = \frac{1}{T} \operatorname*{Min}_{l \in H_1} E_0 \operatorname{Log} \frac{f_0(\mathbf{X}_0, \mathbf{Y}_0) \prod_{t=1}^{T} l_{0t}(X_t, Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{f(\mathbf{X}_0, \mathbf{Y}_0) \prod_{t=1}^{T} l_t(X_t/X_{t-1}^{t-p}, Y_t^{t-p})\, l_t(Y_t/Y_{t-1}^{t-p})}$$

$$= \frac{1}{T} \left\{ \operatorname*{Min}_{f} E_0 \operatorname{Log} \frac{f_0(\mathbf{X}_0, \mathbf{Y}_0)}{f(\mathbf{X}_0, \mathbf{Y}_0)} + \operatorname*{Min}_{l} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(X_t/X_{t-1}^{t-p}, Y_t^{t-p})}{l_t(X_t/X_{t-1}^{t-p}, Y_t^{t-p})} \right.$$

$$\left. + \operatorname*{Min}_{l} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})} \right\}$$

The first two terms are equal to zero since it is possible to set the denominator equal to the numerator. Let us now consider the third term. The optimisation problem appearing in this term can be written:

$$\operatorname*{Max}_{l} \sum_{t=1}^{T} E_0 \operatorname{Log} l_t(Y_t/Y_{t-1}^{t-p})$$

We know, from the computation of D_{13} made in $4.b$ that the solution of this problem is:

$$l_t(y_t/y_{t-1}^{t-p}) = l_{0t}(y_t/y_{t-1}^{t-p})$$

Hence, the discrepancy D_1 is given by:

$$D_1 = \frac{1}{T} \sum_{t=1}^{T} E_0 \operatorname{Log} \frac{l_{0t}(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0t}(Y_t/Y_{t-1}^{t-p})}.$$

b) *Derivation of* D_3

Under the hypothesis of non instantaneous causality H_3, the p. d. f. of $\mathbf{Z}_T$ has the following form:

$$l(\mathbf{z}_T)=f(\mathbf{x}_0, \mathbf{y}_0)\prod_{t=1}^{T} l_t(x_t/X_{t-1}^{t-p}, y_{t-1}^{t-p})\, l_t(y_t/X_{t-1}^{t-p}, y_{t-1}^{t-p})$$

Therefore:

$$D_3=\frac{1}{T}\operatorname*{Min}_{l\in H_3} E_0 \operatorname{Log}\frac{f_0(\mathbf{X}_0), \mathbf{Y}_0)\prod_{t=1}^{T} l_{0\,t}(X_t, Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{f(\mathbf{X}_0, \mathbf{Y}_0)\prod_{t=1}^{T} l_t(X_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})\, l_t(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}$$

$$=\frac{1}{T}\operatorname*{Min}_{f} E_0 \operatorname{Log}\frac{f_0(\mathbf{X}_0, \mathbf{Y}_0)}{f(\mathbf{X}_0, \mathbf{Y}_0)}+\frac{1}{T}\sum_{t=1}^{T} E_0 \operatorname{Log} l_{0\,t}(X_t, Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})$$

$$-\frac{1}{T}\operatorname*{Max}_{l}\sum_{t=1}^{T} E_0 E_0[\operatorname{Log} l_t(X_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})/X_{t-1}^{t-p}, Y_{t-1}^{t-p}]$$

$$-\frac{1}{T}\operatorname*{Max}_{l}\sum_{t=1}^{T} E_0 E_0[\operatorname{Log} l_t(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})/X_{t-1}^{t-p}, Y_{t-1}^{t-p}]$$

It follows from the Kullback inequality that the solutions of the above optimisation problems are given by:

$$f(\mathbf{x}_0, \mathbf{y}_0)=f_0(\mathbf{x}_0, \mathbf{y}_0)$$
$$l_t(x_t/x_{t-1}^{t-p}, y_{t-1}^{t-p})=l_{0\,t}(x_t/x_{t-1}^{t-p}, y_{t-1}^{t-p})$$
$$l_t(y_t/x_{t-1}^{t-p}, y_{t-1}^{t-p})=l_{0\,t}(y_t/x_{t-1}^{t-p}, y_{t-1}^{t-p}).$$

Hence, the discrepancy D_3 is given by:

$$D_3=\frac{1}{T}\sum_{t=1}^{T} E_0 \operatorname{Log}\frac{l_{0\,t}(X_t, Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0\,t}(X_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})\, l_{0\,t}(Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}$$

c) *Derivation of* D_{123}

Since the hypothesis H_{123} is the conditional independence of X_t^1 and Y_t^1 given $\mathbf{Z}_0$ (see 3. *b*), the discrepancy D_{123} is given by the usual formula:

$$D_{123}=\frac{1}{T}\sum_{t=1}^{T} E_0 \operatorname{Log}\frac{l_{0\,t}(X_t, Y_t/X_{t-1}^{t-p}, Y_{t-1}^{t-p})}{l_{0\,t}(X_t/X_{t-1}^{t-p})\, l_{0\,t}(Y_t/Y_{t-1}^{t-p})}.$$

APPENDIX 2

A Property of the Kullback Information Criterion

We prove in this appendix the following general properties of the Kullback information criterion.

Let X, Y, Z be three random vectors whose joint p. d. f., with respect to a given σ-finite measure, is denoted by $f(X, Y, Z)$ we have:

(i) $\mathrm{E}\,\mathrm{Log}\dfrac{f(X/Y, Z)}{f(X/Y)} \geqslant 0$;

(ii) This expectation is equal to zero if and only if: $f(X/Y, Z)=f(X/Y)$ (almost surely) i. e., if and only if X and Z are conditionally independent given Y.

Proof: (i) We know that:

$$\mathrm{E}\ \mathrm{Log}\frac{f(X/Y, Z)}{f(X/Y)} = \mathrm{EE}\left[\mathrm{Log}\frac{f(X/Y, Z)}{f(X/Y)}\middle| Y, Z\right]$$

But the conditional expectation $\mathrm{E}\left[\mathrm{Log}\dfrac{f(X/Y, Z)}{f(X/Y)}\middle| Y, Z\right]$ is always non negative since it can be interpreted as a Kullback information measure.

(ii) The nullity of the expectation

$$\mathrm{E}\,\mathrm{Log}\frac{f(X/Y,Z)}{f(X/Y)}$$

implies the nullity of the Kullback information measure $\mathrm{E}\left[\mathrm{Log}\dfrac{f(X/Y, Z)}{f(X/Y)}\middle| Y, Z\right]$, which implies, from the classical properties of this measure:

$$f(X/Y, Z)=f(X/Y) \quad \text{(almost surely)}.$$

Proof of Theorem 5

Under the stationarity hypothesis H*, we know that:

$$C_{X\to Y}=E_0\,\mathrm{Log}\frac{l_{0\,t}(Y_t(Y_t/X_{t-1}^{t-p},\ Y_{t-1}^{t-p})}{l_{0\,t}(Y_t/Y_{t-1}^{t-p})}$$

does not depend on T.

Therefore, we deduce from theorem 4 and assumption (A) that $CS_{X\to Y}$ converges, as $T\to\infty$, to a limit $CS_{X\to V}^*$ such that:

$$\begin{cases} C_{X\to Y}=CS_{X\to Y}^*+M_Y^* \\ CS_{X\to Y}^*=\lim\limits_{T\to\infty}\dfrac{1}{T}\sum\limits_{t=1}^{T}E_0\,\mathrm{Log}\dfrac{l_{0t}(X_t/X_{t-1}^{1-p},\ Y_T^{1-p}}{l_{0\,t}(X_t/X_{t-1}^{t-p},\ Y_t^{t-p})} \end{cases}$$

From H*, we deduce that

$$E_0\,\mathrm{Log}\ l_{0\,t}(X_t/X_{t-1}^{t-p},\ Y_t^{t-p})$$

does not depend on t.

Hence:

$$CS_{X\to Y}^*=\lim_{T\to\infty}\ A_T-E_0\,\mathrm{Log}\ l_{0\,t}(X_t/X_{t-1}^{t-p},Y_t^{t-p})$$

where:

$$A_T=\frac{1}{T}\sum_{t=1}^{T}a_{t\,T}\qquad\text{with}\quad a_{t\,T}=E_0\,\mathrm{Log}\,l_{0\,t}(X_t/X_{t-1}^{1-p},\ Y_T^{1-p})$$

In order to prove that, for any t, $a_{t\,T}$ converges to lim A_T, as $T\to\infty$, we consider the following steps:

1*st step.* For any T and for any $t=1,2,\ldots,T$ we prove that:

$$a_{t,\,T+1}\geqslant a_{t,\,T}=a_{t+1,\,T+1}$$

2*nd step.* We conclude by noting that the sequence $(a_{1\,T})_{T\in N^*}$ converges.

1*st step.* From the property of the Kullback information criterion proved in Appendix 2, it is clear that for any fixed t.

$$a_{t\,T}=E_0\,\mathrm{Log}\,l_{0\,t}(X_t/X_{t-1}^{1-p},\ Y_T^{1-p})$$

is a nondecreasing function of T.

Moreover, we deduce from the markovian property that:

$$\begin{cases} a_{t\,T}=E_0\,\mathrm{Log}\,l_{0\,t}[X_t/X_{t-1}^{t-p},\ Y_T^{t-p}] \\ a_{t+1,\,T+1}=E_0\,\mathrm{Log}\,l_{0\,t+1}[X_{t+1}/X_t^{t-p+1},\ Y_{T+1}^{t-p+1}] \end{cases}$$

and, using the stationarity assumption H* that:

$$a_{t,\mathrm{T}}=a_{t+1,\mathrm{T}+1}.$$

2nd step:

From the first step we have:

$$a_{t,\mathrm{T}}=a_{t-1,\mathrm{T}-1}=\ldots=a_{t,\mathrm{T}-t+1}$$

Hence:

$$\mathrm{A_T}=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}a_{t\,\mathrm{T}}=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}a_{1,\mathrm{T}-t+1}=\frac{1}{\mathrm{T}}\sum_{\tau=1}^{\mathrm{T}}a_{1,\tau}$$

From the first step $a_{1,\tau}$ is a nondecreasing sequence. The limit of $a_{1,\tau}$ as τ goes to infinity cannot be infinite since the Cesaro mean $\mathrm{A_T}$ converges. Therefore, the sequence $a_{1,\tau}$ converges when T goes to infinity to $\lim\limits_{\mathrm{T}\to\infty}\mathrm{A_T}$.

Since:

$$a_{k\mathrm{T}}=a_{1,\mathrm{T}-k+1}$$

the sequence $(a_{k\mathrm{T}})_{\mathrm{T}\geqslant k}$ converges to the same limit for any fixed k.

In conclusion, each term of the Cesaro mean

$$\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathrm{E_0}\,\mathrm{Log}\frac{l_{0\,t}(\mathrm{X}_t/\mathrm{X}_{t-1}^{t-p},\mathrm{Y}_{\mathrm{T}}^{1-p})}{l_{0\,t}(\mathrm{X}_t/\mathrm{X}_{t-1}^{t-p},\mathrm{Y}_{t}^{t-p})}$$

converges to $\mathrm{CS}^*_{\mathrm{X}\to\mathrm{Y}}$ when T goes to infinity.

A similar result is easily derived for:

$$\mathrm{CS}_{\mathrm{X}\Rightarrow\mathrm{Y}}=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathrm{E_0}\,\mathrm{Log}\frac{l_{0\,t}(\mathrm{X}_t/\mathrm{X}_{t-1}^{t-p},\mathrm{Y}_{\mathrm{T}}^{1-p})}{l_{0\,t}(\mathrm{X}_t/\mathrm{X}_{t-1}^{t-p},\mathrm{Y}_{t-1}^{t-p})}$$

$$=\mathrm{A_T}-\mathrm{E_0}\,\mathrm{Log}\,l_{0\,t}(\mathrm{X}_t/\mathrm{X}_{t-1}^{t-p},\mathrm{Y}_{t-1}^{t-p})$$

since, from H*, $\mathrm{E_0}\,\mathrm{Log}\,l_{0\,t}(\mathrm{X}_t/\mathrm{X}_{t-1}^{t-p},\mathrm{Y}_{t-1}^{t-p})$ does not depend on t.

Each term of the Cesaro mean which defines $\mathrm{CS}_{\mathrm{X}\Rightarrow\mathrm{Y}}$ converges to the same limit $\mathrm{CS}^*_{\mathrm{X}\Rightarrow\mathrm{Y}}$ (when $\mathrm{T}\to\infty$). This limit satisfies:

$$\mathrm{C}_{\mathrm{X}\Rightarrow\mathrm{Y}}=\mathrm{CS}^*_{\mathrm{X}\Rightarrow\mathrm{Y}}+\mathrm{M}^*_{\mathrm{Y}}$$

since

$$\mathrm{C}_{\mathrm{X}\Rightarrow\mathrm{Y}}=\mathrm{CS}_{\mathrm{X}\Rightarrow\mathrm{Y}}+\mathrm{M}_{\mathrm{Y}}$$

and

$\mathrm{C}_{\mathrm{X}\Rightarrow\mathrm{Y}}$ does not depend on t, from H*.

APPENDIX 4

Proof of Theorem 8

We assume that the true p. d. f. belongs to $\tilde{H}$. Then, without loss of generality, we assume that the statistical model is parameterised by a vector of parameters φ which is partitioned into two subvectors $\varphi = (\varphi_1', \varphi_2')'$ such that the hypothesis $\tilde{H}$ is defined by $\varphi_1 = 0$.

With each value $\varphi^0 = (\varphi_1^{0'}, \varphi_2^{0'})'$ of the parameters, we can associate a pseudo-true value $\tilde{\varphi}(\varphi^0) = (0, \tilde{\varphi}_2(\varphi^0)'$, such that $\tilde{\varphi}_2(\varphi^0)$ is solution of the following optimization problem:

$$\underset{\tilde{\varphi}_2}{\mathrm{Min}}\, E_0 \,\mathrm{Log} \frac{l_T(Z_T, \varphi_1^0, \varphi_2^0)}{l_T(Z_T, 0, \tilde{\varphi}_2)}$$

Let us denote by $\hat{\varphi} = (\hat{\varphi}_1, \hat{\varphi}_2')'$ and $\hat{\tilde{\varphi}} = (0, \hat{\tilde{\varphi}}_2')'$ the unconstrained and the constrained maximum likelihood estimators of φ. We have to prove that:

$$\hat{\tilde{D}}^1 - \hat{\tilde{D}}^2 = o_p(N^{-1})$$

where:

$$\hat{\tilde{D}}^1 = \frac{1}{N}[L_N(\hat{\varphi}_1, \hat{\varphi}_2) - L_N(0, \tilde{\varphi}_2(\hat{\varphi}_1, \hat{\varphi}_2))]$$

$$\hat{\tilde{D}}^2 = \frac{1}{N}[L_N(\hat{\varphi}_1, \hat{\varphi}_2) - L_N(0, \hat{\tilde{\varphi}}_2)]$$

1*st step: Expansion of* $\tilde{\varphi}_2$ *around the true value.*

By definition

$$\tilde{\varphi}_2(0, \varphi_2) = \varphi_2 \quad \text{for any } \varphi_2$$

and therefore:

$$\frac{\delta\tilde{\varphi}_2}{\delta\varphi_2'}(0, \varphi_2^0) = \mathrm{Id} \quad \text{(Identity matrix).}$$

Moreover, $\tilde{\varphi}_2(\varphi_1, \varphi_2^0)$ is defined, from the implicit mapping theorem by:

$$k[\varphi_1, \varphi_2^0, \tilde{\varphi}_2(\varphi_1, \varphi_2^0)] = 0$$

$$k[\varphi_1^0, \varphi_2^0, \varphi_2] = E_0 \frac{\delta \,\mathrm{Log}\, l_T(Z_T, 0, \varphi_2)}{\delta\varphi_2}$$

Hence, since $\tilde{\varphi}_2(0, \varphi_2^0) = \varphi_2^0$

$$\frac{\delta\tilde{\varphi}_2}{\delta\varphi_1'}(0, \varphi_2^0) = -\left[\frac{\delta k}{\delta\varphi_2'}(0, \varphi_2^0, \varphi_2^0)\right]^{-1}\left[\frac{\delta k}{\delta\varphi_1'}(0, \varphi_2^0, \varphi_2^0)\right]$$

But

$$\frac{\delta k}{\delta \varphi_2'}(0, \varphi_2^0, \varphi_2^0) = \mathrm{E}_0\left[\frac{\delta^2 \operatorname{Log} l_{\mathrm{T}}(\mathbf{Z}_{\mathrm{T}}, 0, \varphi_2^0)}{\delta \varphi_2 \, \delta \varphi_2'}\right] = -\tilde{\mathrm{I}}_{22}$$

$$\frac{\delta k}{\delta \varphi_1'}(0, \varphi_2^0, \varphi_2^0) = \mathrm{E}_0\left[\frac{\delta \operatorname{Log} l_{\mathrm{T}}(\mathbf{Z}_{\mathrm{T}}, 0, \varphi_2^0)}{\delta \varphi_2} \, \frac{\delta \operatorname{Log} l_{\mathrm{T}}(\mathbf{Z}_{\mathrm{T}}, 0, \varphi_2^0)}{\delta \varphi_1'}\right] = \tilde{\mathrm{I}}_{21}$$

where $\tilde{\mathrm{I}}_{22}$ and $\tilde{\mathrm{I}}_{21}$ are blocks of the Fisher information matrix evaluated for $\varphi = (0, \varphi_2^{0\prime})'$.

Hence:

$$\frac{\delta \tilde{\varphi}_2}{\delta \varphi_1'}(0, \varphi_2^0) = \tilde{\mathrm{I}}_{22}^{-1} \tilde{\mathrm{I}}_{21}$$

Thus, we have:

$$\tilde{\varphi}_2(\varphi_1, \varphi_2) = \tilde{\varphi}_2(0, \varphi_2^0) + \tilde{\mathrm{I}}_{22}^{-1} \tilde{\mathrm{I}}_{21} \, \varphi_1 + \varphi_2 - \varphi_2^0 + o[\| \varphi_1 \| + \| \varphi_2 - \varphi_2^0 \|]$$
$$= \varphi_2^0 + \tilde{\mathrm{I}}_{22}^{-1} (\tilde{\mathrm{I}}_{21}, \tilde{\mathrm{I}}_{22}) \begin{pmatrix} \varphi_1 \\ \varphi_2 - \varphi_2^0 \end{pmatrix} + o[\| \varphi_1 \| + \| \varphi_2 - \varphi_2^0 \|].$$

Hence, if $(0, \varphi_2^{0\prime})$ is the true value of φ, we obtain the following expansion:

$$\tilde{\varphi}_2(\hat{\varphi}_1, \hat{\varphi}_2) = \varphi_2^0 + \tilde{\mathrm{I}}_{22}^{-1} (\tilde{\mathrm{I}}_{21}, \tilde{\mathrm{I}}_{22}) \begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 - \varphi_2^0 \end{pmatrix} + o_p(\mathrm{N}^{-1/2})$$

2nd step: Expansion of $\hat{\tilde{\varphi}}$ around the true value:

As usual, we can consider the likelihood equations

$$\frac{\delta \mathrm{L}_n}{\delta \varphi}(\hat{\varphi}_1, \hat{\varphi}_2) = 0$$

$$\frac{\delta \mathrm{L}_n}{\delta \varphi_2}(0, \hat{\tilde{\varphi}}_2) = 0$$

In order to obtain the following expansions:

$$0 = \frac{1}{\sqrt{\mathrm{N}}} \frac{\delta \mathrm{L}_{\mathrm{N}}}{\delta \varphi_2}(0, \varphi_2^0) - (\tilde{\mathrm{I}}_{21}, \tilde{\mathrm{I}}_{22}) \sqrt{\mathrm{N}} \begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 - \varphi_2^0 \end{pmatrix} + o_p(1)$$

and

$$0 = \frac{1}{\sqrt{\mathrm{N}}} \frac{\delta \mathrm{L}_{\mathrm{N}}}{\delta \varphi_2}(0, \varphi_2^0) - \tilde{\mathrm{I}}_{22} \sqrt{\mathrm{N}} (\hat{\tilde{\varphi}}_2 - \varphi_2^0) + o_p(1)$$

These two expansions imply:

$$\sqrt{N}(\hat{\tilde{\varphi}}_2-\varphi_2^0)=\tilde{I}_{22}^{-1}(\tilde{I}_{21},\tilde{I}_{22})\sqrt{N}\begin{pmatrix}\hat{\varphi}_1\\ \hat{\varphi}_2-\varphi_2^0\end{pmatrix}+o_p(1)$$

$$(\hat{\tilde{\varphi}}_2-\varphi_2^0)=\tilde{I}_{22}^{-1}(\tilde{I}_{21},\tilde{I}_{22})\begin{pmatrix}\hat{\varphi}_1\\ \hat{\varphi}_2-\varphi_2^0\end{pmatrix}+o_p(N^{-1/2}).$$

By comparing the expansions of $\hat{\tilde{\varphi}}_2$ and $\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)$ under $\tilde{H}$, we obtain:

$$\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)-\hat{\tilde{\varphi}}_2=o_p(N^{-1/2})$$

3rd step: Expansions of $L_N(0,\hat{\tilde{\varphi}}_2)$ *and* $L_N(0,\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2))$:

As usual, we have:

$$L_N(0,\hat{\tilde{\varphi}}_2)=L_N(0,\varphi_2^0)+\frac{\delta L_N}{\delta\varphi_2}(0,\varphi_2^0)(\hat{\tilde{\varphi}}_2-\varphi_2^0)$$

$$-\frac{N}{2}(\hat{\tilde{\varphi}}_2-\varphi_2^0)'\tilde{I}_{22}(\hat{\tilde{\varphi}}_2-\varphi_2^0)+o_p(1)$$

Similarly, since we know from the first step that $\sqrt{N}[\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)-\varphi_2^0]$ is asymptotically normally distributed, we have also:

$$L_N[0,\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)]=L_N(0,\varphi_2^0)+\frac{\delta L_N}{\delta\varphi_2}(0,\varphi_2^0)[\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)-\varphi_2^0]$$

$$-\frac{N}{2}[\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)-\varphi_2^0]'\tilde{I}_{22}[\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)-\varphi_2^0]'+o_p(1)$$

These two expansions imply:

$$\begin{aligned}
&L_N(0,\hat{\tilde{\varphi}}_2)-L_N[0,\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)]\\
&\quad=\frac{\delta L_N}{\delta\varphi_2}(0,\varphi_2^0)[\hat{\tilde{\varphi}}_2-\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)]-\frac{N}{2}(\hat{\tilde{\varphi}}_2-\varphi_2^0)'\tilde{I}_{22}(\hat{\tilde{\varphi}}_2-\varphi_2^0)\\
&\qquad+\frac{N}{2}[\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_1)-\varphi_2^0]'\tilde{I}_{22}[\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)-\varphi_2^0]+o_p(1)\\
&\quad=\frac{\delta L_N}{\delta\varphi_2}(0,\varphi_2^0)[\hat{\tilde{\varphi}}_2-\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)].\\
&\qquad+\frac{N}{2}[\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)-\hat{\tilde{\varphi}}_2]'\tilde{I}_{22}[\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)-\hat{\tilde{\varphi}}_2]\\
&\qquad+N(\hat{\tilde{\varphi}}_2-\varphi_2^0)'\tilde{I}_{22}[\tilde{\varphi}_2(\hat{\varphi}_1,\hat{\varphi}_2)-\hat{\tilde{\varphi}}_2]+o_p(1)
\end{aligned}$$

Each term in this last expansion is $o_p(1)$, therefore:

$$L_N(0, \hat{\tilde{\varphi}}_2) - L_N[0, \tilde{\varphi}_2(\hat{\varphi}_1, \hat{\varphi}_2)] = o_p(1)$$

and

$$\hat{\tilde{D}}^1 - \hat{\tilde{D}}^2 = o_p(N^{-1}).$$

Proof of Theorem 11

We have:

$$X^2_{X,Y}(j,l) = X^2_{X \leftrightarrow Y}(j,l) + \sum_{i=1}^{J} \sum_{k=1}^{L} \frac{[p_x(i/j)p_y(k/l) - p(i,./j,l)p(.,k/j,l)]^2}{p_x(l/j)p(.,k/j,l)}$$
$$+2\sum_{i=1}^{J}\sum_{k=1}^{L} \frac{\left\{\begin{array}{r}[p(i,k/j,l)-p(i,./j,l)p(.,k/j,l)][p(i,./j,l)p(.,k/j,l)\\ -p_x(i/j)p_y(k/p]\end{array}\right\}}{p_x(i/j)p(.,k/j,l)}$$

The third term is equal to:

$$2\sum_{i=1}^{J}\frac{p(i,./j,l)}{p_x(i/j)}\sum_{k=1}^{L}[p(i,k/j,l)-p(i,./j,l)p(.,k/j,l)]$$
$$-2\sum_{k=1}^{L}\frac{p_y(k/l)}{p(.,k/j,l)}\sum_{i=1}^{J}[p(i,k/j,l)-p(i,./j,l)p(.,k/j,l)]$$

But:

$$\sum_{k=1}^{L}[p(i,k/j,l)-p(i,./j,l)p(.,k/j,l)] = p(i,./j,l)-p(i,./j,l)=0$$

and

$$\sum_{i=1}^{J}[p(i,k/j,l)-p(i,./j,l)p(.,k/j,l)] = p(.,k/j,l)-p(.,k/j,l)=0.$$

Hence:

$$\chi^2_{X,Y}(j,l) - \chi^2_{X \leftrightarrow Y}(j,l)$$
$$=\sum_{i=1}^{J}\sum_{k=1}^{L}\frac{\left\{\begin{array}{r}[p_x(i/j)p_y(k/l)-p_x(i/j))p(.,k/j,l)+p_x(i/j)p(.,k/j,l)\\ -p(i,./j,l)p(.,k/j,l)]^2\end{array}\right\}}{p_x(i/j)p(.,k/j,l)}$$
$$=\sum_{i=1}^{J}\sum_{k=1}^{L}p_x(i/j)\frac{[p_y(k/l)-p(.,k/j,l)]^2}{p(.,k/j,l).}$$
$$+\sum_{i=1}^{J}\sum_{k=1}^{L}\frac{p(.,k/j,l)}{p_x(i/j)}[p_x(i/j)-p(i,./j,l)]^2$$
$$+2\sum_{i=1}^{J}\sum_{k=1}^{L}[p_y(k/l)-p(.,k/j,l)][p_x(i/j)-p(i,./j,l)].$$

By noting that:

$$\sum_{i=1}^{J} p_x(i/j) = \sum_{i=1}^{J} p(i, . /j, l) = 1$$

$$\sum_{k=1}^{L} p_y(k/l) = \sum_{k=1}^{L} p(., k/j, l) = 1$$

we obtain:

$$\chi^2_{X, Y}(j, l) - \chi^2_{X \leftrightarrow Y}(j, l)$$

$$= \sum_{k=1}^{L} \frac{[p_y(k/l) - p(., k/j, l)]^2}{p(., k/j, l)} + \sum_{i=1}^{J} \frac{[p_x(i/j) - p(i, . /j, l)]^2}{p_x(i/j)}$$

$$= \chi^2_{X \to Y}(j, l) + \chi^2_{X \to Y}(j, l).$$

• References

ANDERSON, T. W. and GOODMAN, L. (1957). — « Statistical Inference about Markov Chains », *AMS*, 28, pp. 89-109.

BOUISSOU, M. B., LAFFONT, J. J. and VUONG Q. H. (1986). — «Tests of Non Causality under Markov Assumptions for Qualitative Panal Data», *Econometrica*, 54, pp. 395-414.

CHAMBERLAIN, G. (1982). — «The General Equivalence of Granger and Sims Causality», *Econometrica*, 50, pp. 569-591.

FLORENS, J. P. and MOUCHART, M. (1982). — «A Note on Non Causality», *Econometrica*, 50, pp. 583-606.

FLORENS, J. P. and MOUCHART, M. (1985). — «A Linear Theory for Non Causality», *Econometrica*, 53, pp. 157-175.

GEWEKE, J. (1982). — «Measurement of Linear Dependence and Feedback between Multiple Time Series», *JASA*, 77, pp. 304-313.

GEWEKE, J. MEESE, R. and DENT, W. (1983). — «Comparing Alternative Tests of Causality in Temporal Systems», *Journal of Econometrics*, 21, pp. 161-194.

GOURIEROUX, C. (1984). — *Économétrie des variables qualitatives*, Economica, Paris.

GRANGER, C. W. J. (1969). — «Investigating Causal Relations by Econometric Models and Cross Spectral Methods», *Econometrica*, 37, pp. 424-439.

GRANGER, C. W. J. (1980). — «Testing for Causality: a Personal Viewpoint». *Journal of Economic Dynamics and Control*, 2, pp. 329-352.

HOSOYA, Y. (1977). — «On the Granger Condition for Non Causality». *Econometrica*, 45, pp. 1735-1736.

PIERCE, D. A. (1979). — «R^2-Measures for Time Series», *JASA*, 74, pp. 901-910.

SIMS, C. A. (1972). — «Money, Income and Causality», *American Economic Review*, 62, pp. 540-552.