

DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

TESTS OF NON-CAUSALITY UNDER MARKOV ASSUMPTIONS
FOR QUALITATIVE PANEL DATA

M. B. Bouissou
Universite des Sciences Sociales, Toulouse

J. J. Laffont
Universite des Sciences Sociales, Toulouse

Q. H. Vuong
California Institute of Technology



SOCIAL SCIENCE WORKING PAPER 501

April 1983
Revised October 1983

ABSTRACT

TESTS OF NON-CAUSALITY UNDER MARKOV ASSUMPTIONS
FOR QUALITATIVE PANEL DATA

by
M.B. Bouissou, J.J. Laffont, and Q.H. Vuong

For many years, social scientists have been interested in obtaining testable definitions of causality (C.W. Granger (1969), C. Sims (1972)). Recent works include those of G. Chamberlain (1982) and J.P. Florens and M. Mouchart (1982). The present paper first clarifies the results of these latter papers by considering a unifying definition of non-causality. Then, log-likelihood ratio (LR) tests for non-causality are derived for qualitative panel data under the minimal assumption that one series is Markov. LR tests for the Markov property are also obtained. Both tests statistics have closed forms. These tests thus provide a readily applicable procedure for testing non-causality on qualitative panel data. Finally, the tests are applied to French Business Survey data in order to test the hypothesis that price changes from period to period are strictly exogenous to disequilibria appearing within periods.

**TESTS OF NON-CAUSALITY UNDER MARKOV ASSUMPTIONS
FOR QUALITATIVE PANEL DATA***

M. B. Bouissou, J. J. Laffont, and Q. H. Vuong

1. Introduction and Summary

For many years, social scientists have been interested in obtaining a testable definition of causality. Earlier contributions include the works of H. A. Simon (1953), R. H. Strotz and H. Wold (1960). Alternative definitions of causality which heavily rely on the stochastic nature of the variables and the principle that the future does not cause the past were then proposed and studied by C. W. Granger (1969) and C. Sims (1972). Recently, G. Chamberlain (1982) and J. P. Florens and M. Mouchart (1982) extended these latter definitions to possibly nonstationary non-gaussian processes. The present paper first clarifies the results of these two recent papers, second, derives some tests for non-causality under minimal assumptions on the process generating the qualitative panel data, and finally, applies the tests to an empirical example.

Throughout the paper, the following definition of non-causality is used: if Y and X are two stochastic processes, then Y does not cause X if at any instant, current and future x 's are independent of past y 's given past x 's. The principal difference between this definition and Granger's definition is that the whole future of X , and not simply its immediate future, must be independent of past y 's given past x 's. By noticing that Granger's definition and

Chamberlain's revised version of Sims' definition are nevertheless both equivalent to the above definition, we reestablish in Section 2, indirectly but in an illuminating way, Chamberlain's general equivalence result.

The essential difficulty in testing for non-causality is that the restrictions imposed by the non-causality of Y on X involves conditioning sets with an infinite number of random variables. To circumvent this difficulty, the X process is assumed to be Markov of a certain order so that the restrictions reduce to independence properties conditional upon finite sets of variables. The restrictions that are imposed on a sample of finite size by the assumptions that X is Markov of a certain order and that Y does not cause X , are derived in Section 3. These restrictions are then decomposed recursively, i.e., in sets of restrictions where each set imposes restrictions on one of the conditional probability distributions of a recursive system.

Using this recursive decomposition, we derive in Section 4, the log-likelihood ratio test of the joint hypothesis that Y does not cause X , and that X is Markov of a certain order when qualitative panel data are available. We also derive the log-likelihood ratio test for a Markov process. It turns out that both test statistics have closed-forms. The two tests therefore provide a readily applicable procedure for testing causality on qualitative variables since no numerical optimization is required. The import of our results is that no assumptions (such as stationarity) on the processes

are made with the exception of the Markov requirement for X. Moreover, by considering qualitative variables, our tests are free of model specification errors since the class of admissible distributions for X and Y need not be a priori restricted.

Our procedure is finally applied to French Business Survey Data in Section 5. The analyzed issue, which is akin to disequilibrium economic theory, involves the relationship between price changes and observed disequilibria on the product market. Specifically, the hypothesis to be tested is whether price changes from period to period is strictly exogenous to intra period disequilibria as measured by some indicator of excess demand or excess supply.

Section 6 contains our conclusion, and an appendix collects proofs of all our theoretical results.

2. Some General Results on Non-Causality

Let X and Y be two possibly non-stationary stochastic scalar or vector processes. In what follows, X and Y are discrete time processes., i.e., $\{(x_t, y_t) : t \text{ in } Z \cup \{-\infty, +\infty\}\}$. Let X_r^s be the set of random variables $\{x_t : r \leq t \leq s\}$. If $r > s$, then X_r^s is by convention the empty set. Similar notations are used for Y.

An important notion for defining non-causality is that of conditional independence. Indeed, if two random variables are conditionally independent given another random variable, then either one of the conditionally independent variables does not provide any

additional information on the other given the knowledge of the conditioning variable. To indicate that the sets of random variables A and B are conditionally independent given the set of random variables C, we use the convenient notation $A \perp B \mid C$.¹

The definitions of non-causality that we consider are those of C. Granger (1969) and C. Sims (1972). More precisely we consider Sims' definition of strict exogeneity as modified by G. Chamberlain (1982). These definitions are:

DEFINITION 1 (Granger Non-Causality): The stochastic process Y does not Granger cause the stochastic process X if and only if

$$(G): \quad x_{t+1} \perp Y_{-\infty}^t \mid X_{-\infty}^t, \text{ for any } t.$$

DEFINITION 2 (Sims-Chamberlain Strict Exogeneity) The stochastic process X is strictly exogenous to the stochastic process Y if and only if

$$(S): \quad X_{t+1}^{\infty} \perp y_t \mid (X_{-\infty}^t, Y_{t-1})$$

for any subset Y_{t-1} of $Y_{-\infty}^{t-1}$ and for any t.

According to Granger's definition, Y does not cause X if, at any instant, the immediate future of X is independent of past and current y's given past and current x's. On the other hand, according to Sims' definition, X is strictly exogenous to Y if, at any instant, current y is independent of future x's given past and current x's and

any past of Y. As is well known, Sims strict exogeneity of X to Y is also a definition of non-causality of Y on X since (S) also states that future x's are independent of current y given current and past x's and any past of Y.

Given that past and current y's may affect some future x's but not the immediate future of X, one may question whether Granger's definition of non-causality is sufficiently strong. This suggests the following definition of non-causality, which we call metaphysical non-causality.

DEFINITION 3 (Metaphysical Non-Causality): The stochastic process Y does not cause the stochastic process X if and only if

$$(C): \quad X_{t+1}^{\infty} \perp Y_{-\infty}^t \mid X_{-\infty}^t, \text{ for any } t.$$

Metaphysical non-causality of Y on X requires that the whole future of X be independent of past and current y's given past and current x's.²

Two remarks are in order. First, the previous definitions apply to completely general discrete-time processes since the X and Y processes need not satisfy any particular assumptions. These definitions can also be extended to continuous-time processes as follows. Let $X_{-\infty}^{t-}$ and X_{t+}^{∞} be respectively the sets of random variables (or σ -fields generated by) $\{x_r : r < t\}$ and $\{x_r : r > t\}$. The set $Y_{-\infty}^{t-}$ is similarly defined. Then the previous definitions apply to continuous-time processes provided "t - 1" and "t + 1" are respectively replaced by "t-" and "t+". Moreover, the results of this

section and the next section can be straightforwardly generalized.

Second, if Y does not Granger cause X at $t = t^0$ only, then we say that (g_{t^0}) holds. The properties (s_{t^0}) and (c_{t^0}) are similarly defined. It is, however, important to note that in order for the stochastic process Y not to cause the stochastic process X according to either one of the above definitions, the corresponding independence restrictions must hold for all t.

It is well known that the (minimum mean square error) linear predictor version of (G) is equivalent to the linear predictor version of (S). (See e.g., C. Sims (1972) for covariance stationary processes with autoregressive representation and no linearly deterministic component, and Y. Hosoya (1977) for more general covariance stationary processes).³ G. Chamberlain (1982), in addition to modifying Sims initial definition, establishes directly the equivalence between (G) and (S).

The remainder of this section provides an indirect but, we think, clarifying proof of G. Chamberlain's general equivalence result (1982, Theorem 4). Our proof is analogous to the one given by R. Kohn (1981) for the linear predictor case with normal processes. We need first some additional definitions and some lemmas. Let $k \geq 1$.

DEFINITION 4 (Granger Non-Causality of order k): The stochastic process Y does not Granger cause, at the order k, the stochastic process X if and only if:

$$(G_k): \quad X_{t+1}^{t+k} \perp Y_{-\infty}^t \mid X_{-\infty}^t, \text{ for any } t.$$

Granger non-causality of order k requires that the k immediate future x 's be jointly independent of past and current y 's given past and current x 's. The next lemma states that (G_k) holds if and only if (G_{k+1}) holds. (Proofs of all stated results can be found in the Appendix.)

LEMMA 1: For any $k \geq 1$, (G_k) is equivalent to (G_{k+1}) .

It follows that Granger non-causality, i.e., (G_1) , is equivalent to any (G_k) .

Granger non-causality of order k involves k future x 's. We can define Sims strict exogeneity of order k by considering current y and $k - 1$ lagged y 's.

DEFINITION 5 (Sims strict Exogeneity of Order k): The stochastic process X is strictly exogenous, at the order k , to the stochastic process Y if and only if :

$$(S_k): \quad X_{t+1}^\infty \perp Y_{t-k+1}^t \mid (X_{-\infty}^t, Y_{t-k}^t);$$

for any subset Y_{t-k}^t of $Y_{-\infty}^{t-k}$, and for any t .

The next result is similar to that of lemma 1. It states that (S_k) holds if and only if (S_{k+1}) holds.

LEMMA 2: For any $k \geq 1$, (S_k) is equivalent to (S_{k+1}) .

Thus, Sims-Chamberlain strict exogeneity, i.e., (S_1) , is equivalent to any (S_k) .

G. Chamberlain's general equivalence result follows from the next theorem as a special case for $k = h = 1$.

THEOREM 1 (General Equivalence Result): For any k and any h , conditions (G_k) , (S_h) , and (C) are all equivalent.

The import of our approach is that (G) and (S) are equivalent because they are both equivalent versions of the same notion which is (C).⁴ Our approach also points out that when (G) holds, i.e., when the immediate future of X is independent of $Y_{-\infty}^t$ given $X_{-\infty}^t$ for any t , then in fact the whole future of X is independent of $Y_{-\infty}^t$ given $X_{-\infty}^t$ for any t . A similar property holds for the strict exogeneity of current and past y 's. It is, however, important to note that these results crucially depend on the requirement that the restrictions associated with (G), (S) or (C) hold for any t .

Of course, there exist equivalent versions of (C) other than (G) and (S). For instance, one may consider the following apparently weaker forms of non-causality of Y on X .

$$(C^*): \quad x_{t+r} \perp Y_{-\infty}^t \mid X_{-\infty}^t, \quad \text{for any } r \geq 1, \text{ and any } t,$$

$$(G_k^*): \quad x_{t+r} \perp Y_{-\infty}^t \mid X_{-\infty}^t, \quad \text{for any } 1 \leq r \leq k, \text{ and any } t.$$

Each of the above conditions is equivalent to (C). Indeed, it is clear that each one is implied by (C). The converse follows from Theorem 1 since each of these conditions implies (G).⁵ It is noteworthy that this latter result implies that (G_k) and (G_k^*) are

actually equivalent.

Finally, G. Chamberlain (1982, Theorem 3) establishes the equivalence between (G) and the following version of Sims strict exogeneity:

$$(S'): X_{t+1}^{\infty} \perp y_t \mid (X_{-\infty}^t, Y_{-\infty}^{t-1}), \text{ for any } t$$

if the following regularity condition (on σ -fields) holds

$$(R): X_{-\infty}^t \cup \left(\bigcap_{k=0}^{+\infty} Y_{-\infty}^{t-k} \right) = X_{-\infty}^t, \text{ for any } t.$$

This result simply follows from Theorem 1 since (S') is equivalent to (C) if (R) holds. Indeed, (C) clearly implies (S'). To see the converse, we note that (S') is equivalent to

$$(S'_k): X_{t+1}^{\infty} \perp Y_{t-k+1}^t \mid (X_{-\infty}^t, Y_{-\infty}^{t-k}) \text{ for any } t \text{ and for any } k \text{ (the proof is similar to that of Lemma 2). Thus (S') implies that}$$

$$X_{t+1}^{\infty} \perp Y_{-\infty}^t \mid (X_{-\infty}^t, \bigcap_{k=0}^{+\infty} Y_{-\infty}^{t-k}) \text{ for any } t \text{ which implies (C) if (R) holds.}$$

3. Non-Causality under Markov Assumptions

The previous section shows that the basic definitions of non-causality, which are Granger's and Sims' definitions, are equivalent to the same general notion which is (C). Thus, from now on, non-causality of Y on X means that the independence restrictions associated with (C) holds.

The essential difficulty in testing for non-causality is that non-causality of Y on X involves a conditioning set with an infinite

number of random variables, namely $X_{-\infty}^t$. Since in general one observes only a finite number of realizations of x's and y's, non-causality of Y on X may not be statistically identified. This follows from the fact that conditional independence between two observed variables given an unobserved variable may not impose any restrictions on the joint probability distribution of the observed variables. For instance, suppose that all the x's and y's are identically null with the exception of x_0, y_1, x_2 where x_0 is unobserved. Suppose that (C) holds so that $x_2 \perp y_1 \mid x_0$. Then (C) may not impose any restrictions on the joint probability distribution of the observed variables (x_2, y_1) .⁶ Hence (C) is not identified.

The previous paragraph points out that one needs to introduce additional assumptions on the X and Y processes in order to test for non-causality of Y on X. To circumvent the problem of conditioning on sets with infinite number of variables, one may simply assume that the X process starts at $t = 1$ (the first period of the sample), or equivalently that the values of x's prior to $t = 1$ are identically null. It is clear that such an assumption does not correspond to most economic time series. Then, one may instead assume that the X and Y processes are jointly stationary, as it is usually done in econometric works.⁷ It can, however, be shown, by modifying the example given in the previous paragraph, that the stationarity assumption is not always sufficient to ensure that (C) is identified. Thus, one must in addition a priori restrict the class of probability distributions to be considered, i.e., one must specify the probability model generating

the stationary processes X and Y . It follows that the inference that one can make about non-causality is conditional upon the truthfulness of the additional assumptions that one put forward to identify (C).

Since the question of whether any statement can be made about non-causality based just on statistical data is important, as C. Granger (1980) argued, it is essential that one invokes additional assumptions on the X and Y processes that are relatively weak and easily testable. The only additional assumption that is used in the present paper is that the stochastic process X is Markov of some order. In particular, the X and Y processes need not be stationary. Moreover, the Y process need not be Markov. This is simply because we are testing for the non-causality of Y on X . Finally, it is important to note that we do not actually require the formulation of a probability model for the X and Y processes so that our tests derived thereafter are necessarily free of any specification errors.

In this section, we first derive the restrictions that are imposed on the stochastic processes X and Y when Y does not cause X and X is Markov of some order. Then, we consider the maximum number of restrictions that are imposed on a sample of finite size by the non-causality of Y on X and the Markov requirement on X .

Let m be an integer possibly equal to zero.⁸ By a Markov process of order m , we mean the following:

DEFINITION 6 (Markov Process of order m): The stochastic process X is Markov of order m if and only if:

$$(M_m): X_{t+1}^\infty \perp X_{-\infty}^{t-m} \mid X_{t-m+1}^t, \text{ for any } t.$$

In words, the stochastic process X is Markov of order m if and only if, at any instant, the future of X is independent of the past of X given current and $m-1$ lagged x 's. As is well known, the stochastic process X is Markov of order m if and only if it is an autoregressive process of order m , i.e., an $AR(m)$:

$$AR(m): x_{t+1} \perp X_{-\infty}^{t-m} \mid X_{t-m+1}^t, \text{ for any } t.^9$$

The next lemma determines the set of independence restrictions imposed on the stochastic processes X and Y when Y does not cause X and X is Markov of order m .

LEMMA 3: For any $m \geq 0$, (C) and (M_m) both hold if and only if (R_m) holds, where

$$(R_m): X_{t+1}^\infty \perp (X_{-\infty}^{t-m}, Y_{-\infty}^t) \mid X_{t-m+1}^t, \text{ for any } t.$$

Condition (R_m) requires that, at any time, the future of X is independent of past x 's and current and past y 's given the m most recent x 's. It is clear that the principal use of the Markov assumption on the X process is to replace the independence restrictions associated with (C) by independence restrictions that now involve only finite sets of conditioning variables.

In most situations, one does not observe the X and Y processes over the whole time axis, but only for finite number of periods. Let

$t = 1$ be the beginning of the sampling period, and T be the number of periods for which the X and Y processes are observed. It is now possible to derive the restrictions that are implied by (R_m) on the joint probability distribution of the observed variables (X_1^T, Y_1^T) .

Since we shall eventually be interested in testing the validity of our additional assumption that X is Markov, we begin with the restrictions implied by (M_m) . From now on, we assume that $T \geq m+2$. Indeed, if this were not the case, we would not be able to test whether or not X is Markov of order m since the restrictions (M_m) would not be identified. Then, it is straightforward to see that the restrictions implied by (M_m) on the joint probability distribution of X_1^T are:

$$(M_m^T): X_{t+1}^t \perp X_1^{t-m} \mid X_{t-m+1}^t, \text{ for any } t = m+1, \dots, T-1.$$

These are all the possible restrictions implied by (M_m) alone since no observations are available prior to time 1 and after time T . It is worth noting that each restriction of (M_m^T) involves a conditioning set of variables that are all observed.

We now turn to the restrictions implied by the non-causality of Y on X and the Markov assumption on X . It can readily be seen that these restrictions are:

$$(R_m^T): X_{t+1}^T \perp (X_1^{t-m}, Y_1^t) \mid X_{t-m+1}^t, \text{ for any } t = m, \dots, T-1.$$

As before, these are all the possible restrictions implied by (R_m) alone on the joint probability distribution of the observed variables

(X_1^T, Y_1^T) . Moreover, as for (M_m^T) , each restriction of (R_m^T) involves a conditioning set of only observed variables.¹⁰ It is worth noting that the problem of initial conditions has been avoided. This actually follows from our desire of obtaining results under minimal assumptions.

The next theorem presents the basic result that underlies the tests for non-causality derived in the next section. It essentially provides a recursive decomposition of the $T - m$ restrictions of (R_m^T) .

THEOREM 2 (A recursive Decomposition of (R_m^T)): For any $m \geq 0$, (R_m^T) holds if and only if the following conditions simultaneously hold:

- (i) (M_m^T) , and
- (ii) $(c_m^T): X_{m+1}^T \perp Y_1^m \mid X_1^m$,
- (iii) for every $t = m+1, \dots, T-1: (s_t^T): X_{t+1}^T \perp y_t \mid (X_1^t, Y_1^{t-1})$.

Condition (i) simply requires that the restrictions on the joint distribution of X_1^T that are implied by the Markov assumption on X hold. Hence the probability model for the observed variables (X_1^T, Y_1^T) that is associated with the restrictions (R_m^T) is nested in the probability model associated with the restrictions (M_m^T) . Condition (ii) is simply condition (C) written for only one period (namely $t = m$, which is the first period for which one observes $m - 1$ lagged x 's) as if the x process was starting at $t = 1$. Similarly, for any $t \geq m + 1$, each condition (s_t^T) is Sims condition written at t only, as if the X and Y processes were both starting at $t = 1$.¹¹

The import of Theorem 2 is to provide a convenient way to impose the various restrictions of (R_m^T) . Specifically, condition (M_m^T) bears only on the observed x 's. Condition (ii) can be interpreted as stating that the variables Y_1^m are independent of the variables X_{m+1}^1 conditional upon all the other observed x 's. Condition (iii) means that, for any $t \geq m+1$, y_t is independent of the variables X_{t+1}^T conditionally upon all the observed x 's and all the previous observed y 's. Since " $A \perp B \mid C$ " is equivalent to the non-dependence on B of the conditional probability distribution of A given (B,C) , it follows that the restrictions imposed by (R_m^T) on the joint probability distribution $\Pr(X_1^T, Y_1^T)$ of the observed variables can readily be specified by considering the recursive system of joint and conditional probability distributions, $\Pr(X_1^T)$, $\Pr(Y_1^m \mid X_1^T)$ and $\Pr(y_t \mid X_1^T, Y_1^{t-1})$ for $t = m+1, \dots, T-1$.

4. Tests of Non-Causality under Markov Assumptions

If one does not invoke any additional assumptions, such as stationarity, one requires panel data in order to estimate a model. Indeed, panel data allows one to observe many realizations of the X and Y processes. Moreover, if one does not want to a priori restrict, by further distributional assumptions, the class of probability distribution $\Pr(X_1^T, Y_1^T)$ that satisfy (R_m^T) , then the easiest way to proceed is to consider qualitative data. This is so because, with qualitative data, one has available non-parametric tests based on goodness-of-fit statistics such as Pearson chi-square statistics and

log-likelihood Ratio (LR) statistics (see e.g. L.A. Goodman (1978), S.J. Haberman (1974)), that can be used to test a model directly against the set of all possible probability distributions, i.e., against the so-called saturated model.

From now on, it is assumed that one observes n independent realizations of the $2T$ random variables (X_1^T, Y_1^T) . Moreover, for any $t = 1, \dots, T$, it is assumed that x_t and y_t are qualitative random variables with I_t and J_t categories respectively.¹² The indices i_t and j_t are used to indicate the values taken on by x_t and y_t .

In the previous section, we have derived the restrictions that are imposed on the observed random variables by the non-causality of Y on X and the assumption that X is Markov. Since, for any m , the restrictions (R_m^T) do not involve the variable y_T , we shall consider the restrictions imposed on the joint probability distribution $\Pr(X_1^T, Y_1^{T-1})$. For any $i_1^T = (i_1, \dots, i_T)$ and $j_1^{T-1} = (j_1, \dots, j_{T-1})$, we let $p(i_1^T, j_1^{T-1})$ be the probability that X_1^T and Y_1^{T-1} are respectively equal to i_1^T and j_1^{T-1} . More generally, $p(i_x^s, j_t^u)$ denotes the probability that X_x^s and Y_t^u are respectively equal to i_x^s and j_t^u .

Since the n realizations of the X and Y processes are independent and since all the variables are qualitative, the contingency table associated with (X_1^T, Y_1^{T-1}) is a sufficient statistics. This contingency table is simply the vector $\{n(i_1^T, j_1^{T-1})\}$, for any (i_1^T, j_1^{T-1}) where $n(i_1^T, j_1^{T-1})$ is the number of observations such that $X_1^T = i_1^T$ and $Y_1^{T-1} = j_1^{T-1}$. The marginal contingency table $\{n(i_x^s, j_t^u)\}$, for any (i_x^s, j_t^u) is similarly defined with respect to the

subset of variables (X_r^s, Y_t^u) . The marginal contingency table is readily obtained from the full contingency table by simply adding up the $n(i_1^T, j_1^{T-1})$'s over the indices that are not associated with the variables of the subset.

Since non-causality of Y on X is identified only under additional assumptions, we shall first solve the problem of testing the Markov assumption on X. Since this latter assumption bears only on X_1^T , we can simply consider the joint probability distribution of X_1^T . The log-likelihood is:

$$\log L_o = \sum_{i_1^T} n(i_1^T) \log p(i_1^T). \quad (4.1)$$

In order to derive the LR-test of the hypothesis that X is Markov of order m, it is necessary to maximize the log-likelihood under the restrictions (M_m^T) . The next lemma gives the Maximum-Likelihood (M.L.) estimates of the probabilities $p(i_1^T)$ under the restrictions (M_m^T) . The import of the result is that the M.L. estimates have a closed form so that they can readily be computed. The lemma simply used the fact that the set of strictly positive probability distributions $\Pr(X_1^T)$ that satisfy (M_m^T) is a joint log-linear probability model for X_1^T .¹³

LEMMA 4: For any $m \geq 0$ and for any i_1^T , the M. L. estimate of $p(i_1^T)$ under the restrictions (M_m^T) is:

$$\hat{p}^m(i_1^T) = \frac{\prod_{t=1}^{T-m} n(i_t^{t+m})}{n \prod_{t=1}^{T-m-1} n(i_{t+1}^{t+m})}. \quad (4.2)$$

The convention $0 \div 0 = 0$ is used in the above lemma and in the next results.¹⁴

It is now straightforward to obtain the LR statistic for testing the hypothesis that X is Markov of order m against the hypothesis of no restrictions on X. Let

$$LR_o^m = 2 \sum_{i_1^T} n(i_1^T) \log \frac{n(i_1^T)}{\hat{p}^m(i_1^T)}. \quad (4.3)$$

The next result essentially gives the number of degrees of freedom of the LR statistic.

THEOREM 3 (LR Test for a Markov of Order m): For any m such that $0 \leq m \leq T - 2$, LR_o^m is the LR statistic for testing the null hypothesis that X is Markov of order m against the hypothesis of no restrictions on X. For large n, and under the null hypothesis, this statistic follows a chi-square distribution with number of degrees of freedom

$$ddf_o^m = \left(\prod_{t=1}^T I_t \right) - \left[\sum_{t=1}^{T-m} \left(\prod_{k=t}^{t+m} I_k \right) - \sum_{t=1}^{T-m-1} \left(\prod_{k=t+1}^{t+m} I_k \right) \right] \quad (4.4)$$

As a consequence of Theorem 3, it is possible to test the hypothesis that X is Markov of order m against the hypothesis that X is Markov of order r where $r \geq m+1$. The first hypothesis is clearly nested in the latter hypothesis since if X is Markov of order m then X is necessarily Markov of order r for any $r \geq m+1$. For identification of the maintained hypothesis, it is assumed that $r \leq T+2$. Let

$$LR_r^m = 2 \sum_{i_1^T} n(i_1^T) \log \frac{\hat{p}^r(i_1^T)}{\hat{p}^m(i_1^T)}. \quad (4.5)$$

where $\hat{p}^r(i_1^T)$ is the M.L. estimate of $p(i_1^T)$ under the restrictions (M_r^T) .¹⁵

COROLLARY 1: For any (m, r) such that $1 \leq m+1 \leq r \leq T-2$, LR_r^m is the LR statistic for testing the null hypothesis that X is Markov of order m against the alternative hypothesis that X is Markov of order r . For n large, and under the null hypothesis, this statistic follows a chi-square distribution with number of degrees of freedom

$$ddf_r^m = ddf_0^r - ddf_0^m, \quad (4.6)$$

where ddf_0^r and ddf_0^m are given by (4.4).

We now turn to the testing of the non-causality of Y on X given the maintained hypothesis that X is Markov of order m . As noted in Section 2, Theorem 2 gives a recursive decomposition of the restriction (R_m^T) . Specifically, since

$$\Pr(X_1^T, Y_1^{T-1}) = \Pr(X_1^T) \cdot \Pr(Y_1^m | X_1^T) \cdot \prod_{t=m+1}^{T-1} \Pr(y_t | X_1^T, Y_1^{t-1}) \quad (4.7)$$

it follows that, instead of considering the set of distributions $\Pr(X_1^T, Y_1^{T-1})$ that satisfy (R_m^T) , we can equivalently consider the recursive system of probability models in which (i) $\Pr(X_1^T)$ satisfies the restrictions (M_m^T) , (ii) $\Pr(Y_1^m | X_1^T)$ satisfies (c_m^T) , and (iii) for

every $t = m+1, \dots, T-1$, $\Pr(y_t | X_1^T, Y_1^{t-1})$ satisfies (s_t^T) .

Moreover, the log-likelihood function associated with the observed variables (X_1^T, Y_1^{T-1}) is:

$$\begin{aligned} \text{Log } L &= \sum_{(i_1^T, j_1^{T-1})} n(i_1^T, j_1^{T-1}) \log p(i_1^T, j_1^{T-1}) \\ &= \text{Log } L_0 + \text{Log } L_m + \sum_{t=m+1}^{T-1} \text{Log } L_t \end{aligned} \quad (4.8)$$

where $\text{Log } L_0$ is given by (4.1), and

$$\text{Log } L_m = \sum_{(i_1^T, j_1^m)} n(i_1^T, j_1^m) \log p(j_1^m | i_1^T) \quad (4.9)$$

$$\text{Log } L_t = \sum_{(i_1^T, j_1^t)} n(i_1^t, j_1^t) \log p(j_t | i_1^T, j_1^{t-1}) \quad (4.10)$$

for any $t = m+1, \dots, T$. Hence the log-likelihood function $\text{Log } L$ is simply the sum of the marginal and conditional log-likelihood functions associated with the probability models composing the recursive system. As a matter of fact, this system is a recursive system of Conditional Log-Linear Probability (CLLP) models (see Q. H. Vuong (1982)). It follows that the M. L. estimation of the joint probability distribution $\Pr(X_1^T, Y_1^{T-1})$, under the restrictions (R_m^T) , can readily be obtained from (4.7) by estimating separately each of the probability models of the recursive system by the maximum-likelihood method.¹⁶

The next lemma gives the (conditional) M.L. estimates of $\Pr(Y_t^m \mid X_1^T)$ under the restrictions (c_m^T) , and of $\Pr(Y_t \mid X_1^T, Y_1^{t-1})$ under the restrictions (s_t^T) . As for lemma 4, the import of the result is that the M. L. estimates have a closed form and hence are readily computed.

LEMMA 5: For any $m \geq 0$ and for any (i_1^T, j_1^m) , the (conditional) M. L. estimate of $p(j_1^m \mid i_1^T)$ under the restrictions (c_m^T) is

$$\hat{p}(j_1^m \mid i_1^T) = \frac{n(i_1^m, j_1^m)}{n(i_1^m)}, \quad (4.11)$$

and for any $t = m+1, \dots, T-1$ and for any (i_1^T, j_1^T) , the (conditional) M. L. estimate of $p(j_t \mid i_1^T, j_1^{t-1})$ is

$$\hat{p}(j_t \mid i_1^T, j_1^{t-1}) = \frac{n(i_1^t, j_1^t)}{n(i_1^t, j_1^{t-1})}. \quad (4.12)$$

From (4.8)-(4.12), we can readily derive the LR statistics for testing the joint hypothesis that Y does not cause X and X is Markov of order m, against the hypothesis of no restrictions on X and Y. Let

$$LR_{c+m} = LR_0^m + LR_m^m + \sum_{t=m+1}^{T-1} LR_t^m \quad (4.13)$$

where LR_0^m is given by (4.3), and

$$LR_m^m = 2 \sum_{(i_1^T, j_1^m)} n(i_1^T, j_1^m) \log \left[\frac{n(i_1^T, j_1^m)}{n(i_1^T)} \cdot \frac{n(i_1^m)}{n(i_1^m, j_1^m)} \right] \quad (4.14)$$

$$LR_t^m = 2 \sum_{(i_1^T, j_1^t)} n(i_1^T, j_1^{T-1}) \log \left[\frac{n(i_1^T, j_1^t)}{n(i_1^T, j_1^{t-1})} \cdot \frac{n(i_1^t, j_1^{t-1})}{n(i_1^t, j_1^t)} \right] \quad (4.15)$$

for any $t = m+1, \dots, T-1$. The next result essentially gives the formula for the number of degrees of freedom of the LR statistic.

THEOREM 4 (LR Test for Non-Causality and Markov of Order m): For any m such that $0 \leq m \leq T-2$, LR_{c+m} is the LR statistic for testing the null hypothesis that Y does not cause X and that X is Markov of order m against the hypothesis of no restrictions on X and Y. For large n and under the null hypothesis, this statistic follows a chi-square distribution with number of degrees of freedom

$$ddf_{c+m} = ddf_0^m + ddf_m^m + \sum_{t=m+1}^{T-1} ddf_t^m \quad (4.16)$$

where ddf_0^m is given by (4.4), and

$$ddf_m^m = \left[\left(\prod_{k=1}^m J_k \right) - 1 \right] \left[\left(\prod_{h=1}^T I_h \right) - \left(\prod_{h=1}^k I_k \right) \right] \quad (4.17)$$

$$ddf_t^m = (J_t - 1) \left[\prod_{h=1}^T I_h \prod_{k=1}^{t-1} J_k - \prod_{h=1}^t I_h \prod_{k=1}^{t-1} J_k \right] \quad (4.18)$$

for any $t = m+1, \dots, T-1$.

The statistic LR_{c+m} is used to test the joint hypothesis that Y does not cause X and that X is Markov of order m against the hypothesis of no restrictions on X and Y. One may also want to test that Y does not cause X under the maintained hypothesis that X is

Markov of order m . Let

$$LR_c^m = LR_m^m + \sum_{t=m+1}^{T-1} LR_t^m \quad (4.19)$$

where LR_m^m and LR_t^m are respectively given by (4.14) and (4.15). The next result is an immediate corollary of Theorem 4.

COROLLARY 2 (LR Test for Non-Causality under Markov of Order m): For any m such that $0 \leq m \leq T-2$, LR_c^m is the LR statistic for testing the null hypothesis that Y does not cause X and X is Markov of order m against the maintained hypothesis that X is Markov of order m . For large n , and under the null hypothesis, this statistic follows a chi-square distribution with number of degrees of freedom.

$$ddf_c^m = ddf_m^m + \sum_{t=m+1}^{T-1} ddf_t^m \quad (4.20)$$

where ddf_m^m and ddf_t^m are respectively given by (4.17) and (4.18).

It is worth noting that we can also separately test each of the sets of restrictions $(c_m^T), (s_{m+1}^T), \dots, (s_{T-1}^T)$ that are imposed by the non-causality of Y on X under the maintained hypothesis that X is Markov of order m . Specifically, from Corollary 1, the sets of restrictions (c_m^T) and (s_t^T) can be separately tested under (M_m^T) by using respectively the statistics LR_m^m and LR_t^m that are given by (4.14) and (4.15). The degrees of freedom of these statistics are respectively ddf_m^m and ddf_t^m as defined by (4.17) and (4.18).

5. An Empirical Application

Since the initial theoretical work in disequilibrium economics of R. Barro and H. Grossman (1978), J. P. Benassy (1982), and E. Malinvaud (1977), fix-price models have been estimated frequently (see J. J. Laffont (1983) for a survey of recent empirical work). The fix-price paradigm does not, however, imply that prices never change:

"... we do not mean that prices will remain the same in the period under study as they did in the preceding period; we simply mean that their movement is 'autonomous': it is not significantly influenced for our purpose by the formation of demands and supplies on which attention will concentrate."

(E. Malinvaud (1977, p.12))

The purpose of this section is to test that price movement is indeed autonomous. Specifically, we shall test whether price changes from period to period are not caused by disequilibria appearing within previous periods. As seen in Section 2, this is equivalent to testing that price changes from period to period are strictly exogenous to intra-period disequilibria. Then we shall test whether price changes from period to period are not caused by current and past disequilibria.

The data that we use has been collected by the Institut National de la Statistique et des Etudes Economiques (INSEE) from about 4000 firms through periodic Business Survey Tests taken each year in March, June, and November, starting from June 74 to November 78.¹⁷ We shall be interested in the disequilibrium experienced by

each firm on its good market.¹⁸ Let ID be the indicator of the type of disequilibrium. This variable is dichotomous and is constructed from the answer to the question:

"If you receive more orders could you produce more with your actual capacities?"

If the firm answers YES we presume that there is excess supply (ID=1), while if the firm answers NO we presume that there is excess demand (ID=2).¹⁹

Let IP be the indicator of the price change from period to period. This variable is trichotomous and is constructed from the answer to the question:

"Would you indicate the variation of your sales prices (net of tax) since the last survey?"

The first category, IP=1, is constructed so that it corresponds to an increase in real terms; the second category, IP=2, to a stability; and the third category, IP=3, to a decrease.²⁰

Our first problem is to know whether the price variations IP are strictly exogenous to the disequilibrium indicator ID. Hence we test the null hypothesis that ID does not cause IP. As discussed in the previous sections, we first need to accept a Markov of some order on the IP process. We have then restricted our analysis to the consumption good sector.²¹ The average number of respondents over three successive surveys drops, however, to about 400. Given that the dimension of the contingency table for testing noncausality of ID on

IP for a series of three successive periods is already $3^3 \times 2^2$, i.e. 108, we could at most test a Markov of order 1 on IP (see footnote 21).

Table 1 presents our results when analyzing three successive surveys.²² The first column indicates the date of the third survey; the second column gives the number of firms for which observations on ID and IP are available for the corresponding three surveys; the third column gives the LR statistic (4.3) for T=3 which is used to test the hypothesis that the IP process is Markov of order 1; the fourth column gives the LR statistic (4.19) for T=3 and m=1 which is used to test the hypothesis that ID does not cause IP given that IP is Markov of order 1; finally the fifth column gives the LR statistic (4.13) for T=3 and m=1 which is used to test the joint hypothesis that ID does not cause IP and that IP is Markov of order 1.

Our results show that we cannot reject at the 10% significance level the hypothesis that the IP process is Markov of order 1 for 6 out of 11 periods. For these 6 periods, the hypothesis that ID does not cause IP cannot be rejected at the 10% level. Our results thus support the hypothesis that changes in prices from period to period are strictly exogenous to the disequilibria appearing within periods.

TABLE 1
LR Statistics with
Upper-Tail Probabilities in parentheses

Ending Periods TE = 3	Number of Cases	For Markov of Order 1 on X DF = 12	For Non-Causality of Y on X assuming Markov of Order 1 on X DF = 60	For Non-Causality of Y on X and for Markov of Order 1 on X DF = 72
75-03	413	12.5 * (40.8)	59.2 * (50.4)	71.7 * (48.8)
75-06	397	16.5 * (17.1)	37.7 * (98.9)	54.1 * (94.2)
75-11	386	30.5 (.002)	30.8 * (99.9)	61.3 * (81.1)
76-03	387	12.6 * (39.8)	60.2 * (46.7)	72.8 * (45.0)
76-06	398	32.8 (.001)	68.9 * (20.2)	101.7 (.012)
76-11	384	52.1 (.000)	72.2 * (13.4)	124.2 (.000)
77-03	345	8.9 * (71.2)	68.9 * (20.2)	77.8 * (30.0)
77-06	356	13.4 * (33.9)	59.3 * (50.0)	72.7 * (45.2)
77-11	395	29.2 (.004)	74.3 * (10.2)	103.5 (.009)
78-03	367	16.1 * (18.5)	65.1 * (30.4)	81.1 * (21.5)
78-11	401	31.6 (.002)	62.2 * (39.9)	93.7 (.044)

* indicates that the null hypothesis cannot be rejected at the 10% significance level.

The previous results use Definition 3 of non-causality which states that ID does not cause IP if and only if $IP_{t+1}^{\infty} \perp ID_{-\infty}^t \mid IP_{-\infty}^t$ for any t. One may wonder whether our qualitative results would still hold if one also includes the current realization of ID, i.e., ID_{t+1} . This leads to the following revised definition of non-causality which we call E-non-causality where E stands for extended.

DEFINITION 7 (Extended Non-Causality): The stochastic process Y does not E-cause the stochastic process X if and only if

$$(EC): X_{t+1}^{\infty} \perp Y_{-\infty}^{t+1} \mid X_{-\infty}^t \text{ for any } t.$$

It is clear that Y does not E-cause X if and only if,

according to definition 3, \tilde{Y} does not cause X, where $\tilde{Y}_t = Y_{t+1}$ for any t. It follows that we can use the LR statistics derived earlier to test that ID does not E-cause IP.²³

Table 2 displays the corresponding statistics. As can readily be seen, the results are quite similar to those of Table 1. On the whole, our data supports the hypothesis that price changes are not caused by current and past disequilibria.

TABLE 2
LR Statistics with
Upper-Tail Probabilities in parentheses

Finding Periods	Number of Cases	For Markov of	For E-Non-Causality	For E-Non-Causality
		Order 1 on X DF = 12	of Y on X assuming Markov of Order 1 on X DF = 60	of Y on X and for Markov of Order 1 on X DF = 72
75-03	393	11.5 * (48.8)	59.1 * (51.0)	70.6 * (52.6)
75-06	369	15.1 * (23.5)	36.1 * (99.4)	51.2 * (97.0)
75-11	373	33.1 (.001)	37.8 * (98.9)	70.9 * (51.4)
76-03	390	13.4 * (34.2)	71.9 * (13.9)	85.3 * (13.5)
76-06	388	40.7 (.000)	45.7 * (91.4)	86.4 * (11.8)
76-11	374	53.4 (.000)	56.8 * (59.4)	110 (.003)
77-03	354	5.7 * (93.3)	64.7 * (31.8)	70.3 * (53.5)
77-06	353	13.0 * (36.6)	63.6 * (35.2)	76.6 * (33.3)
77-11	397	33.7 (.001)	69.5 * (18.8)	103.2 (.009)
78-03	367	12.3 * (42.4)	51.0 * (79.0)	63.3 * (76.0)
78-11	404	35.0 (.000)	53.4 * (71.6)	88.4 (.092)

* indicates that the null hypothesis cannot be rejected at the 10% significance level.

6. Conclusion

In this paper, we have introduced a unifying definition of non-causality which was proved to be equivalent to Granger's definition of non-causality and to Chamberlain's revised version of Sims' strict exogeneity.

After having argued that non-causality of Y on X is by itself non-identified in practice, we have introduced the additional assumption that X is Markov of some order. Then, using a recursive decomposition of all the restrictions that are imposed on a panel data by the non-causality of Y on X and the Markov assumption on X, we have derived the log-likelihood ratio tests for testing the following three hypotheses: (i) X is Markov of order m, (ii) Y does not cause X given that X is Markov of order m, and (iii) Y does not cause X and that X is Markov of order m.

It turns out that all the test statistics have closed-forms. These tests therefore provide a readily applicable procedure for testing non-causality on qualitative panel data. Moreover, these tests are free of model specification errors since the form of the relationship between Y and X need not be a priori specified.

Finally, the procedure is applied to French Business Survey data to test the hypothesis that price changes from period to period are strictly exogenous to intra-period disequilibria as measured by an indicator of excess demand or excess supply. Our empirical results show that this hypothesis, which is crucial to the relevance of disequilibrium economics, cannot be rejected at the 10% significance level.

APPENDIX

I. The following fundamental property of conditional independence (FPCI) is used to prove the results of Sections 2 and 3. Let A, B, C, D, be 4 sets of random variables. Then $A \perp (B,C) \mid D$ if and only if

$$(i) \quad A \perp B \mid (C,D) \quad \text{and}$$

$$(ii) \quad A \perp C \mid D$$

(see, e.g., J. P. Florens and M. Mouchart (1982, Theorem A.1, p. 588))

PROOF OF LEMMA 1: (G_{k+1}) implies (G_k) . To prove the converse, it suffices to write (G_k) at $t+1$:

$$X_{t+2}^{t+k+1} \perp Y_{-\infty}^{t+1} \mid X_{-\infty}^{t+1}, \text{ for any } t$$

which implies

$$X_{t+2}^{t+k+1} \perp Y_{-\infty}^t \mid X_{-\infty}^{t+1}, \text{ for any } t. \quad (A.1)$$

On the other hand, (G_k) implies $(G_1) = (G)$ so that:

$$X_{t+1} \perp Y_{-\infty}^t \mid X_{-\infty}^t, \text{ for any } t. \quad (A.2)$$

From (A.1), (A.2), and the FPCI, it follows that

$$X_{t+1}^{t+k+1} \perp Y_{-\infty}^t \mid X_{-\infty}^t, \text{ for any } t.$$

Q.E.D.

PROOF OF LEMMA 2: (S_k) obviously implies (S_{k+1}) . Let Y_{t-k-1} be a

subset of $Y_{-\infty}^{t-k-1}$. Since $Y_{t-k-1} \cup y_{t-k}$ is a subset of $Y_{-\infty}^{t-k}$, and since (S_k) holds at t , we have:

$$X_{t+1}^{\infty} \perp Y_{t-k+1}^t \mid (X_{-\infty}^t, Y_{t-k-1}, y_{t-k}), \text{ for any } t$$

which implies from the FPCI:

$$X_{t+1}^{\infty} \perp y_t \mid (X_{-\infty}^t, Y_{t-k-1}, Y_{t-k}^{t-1}), \text{ for any } t. \quad (A.3)$$

Let us now write (S_k) at $t-1$ for the subset Y_{t-k-1} of $Y_{-\infty}^{t-k-1}$:

$$X_t^{\infty} \perp Y_{t-k}^{t-1} \mid (X_{-\infty}^{t-1}, Y_{t-k-1}), \text{ for any } t. \quad (A.4)$$

From (A.3), (A.4), and the FPCI, it follows that:

$$X_{t+1}^{\infty} \perp Y_{t-k}^t \mid (X_{-\infty}^t, Y_{t-k-1}), \text{ for any } t,$$

i.e., (S_{k+1}) .

To prove that (S_{k+1}) implies (S_k) , we consider 2 cases. (i)

Suppose that Y_{t-k} does not contain y_{t-k} . Then Y_{t-k} is a subset of $Y_{-\infty}^{t-k-1}$ so that from (S_{k+1}) we get:

$$X_{t+1}^{\infty} \perp Y_{t-k}^t \mid (X_{-\infty}^t, Y_{t-k}), \text{ for any } t,$$

which implies (S_k) , i.e.:

$$X_{t+1}^{\infty} \perp Y_{t-k+1}^t \mid (X_{-\infty}^t, Y_{t-k}), \text{ for any } t.$$

(ii) Suppose that Y_{t-k} does contain y_{t-k} . Then $Y_{t-k} = y_{t-k} \cup Y_{t-k-1}$ where Y_{t-k-1} is a subset of $Y_{-\infty}^{t-k-1}$. From (S_{k+1}) it follows that:

$$X_{t+1}^{\infty} \perp Y_{t-k}^t \mid (X_{-\infty}^t, Y_{t-k-1}^t), \text{ for any } t,$$

which implies:

$$X_{t+1}^{\infty} \perp Y_{t-k+1}^t \mid (X_{-\infty}^t, Y_{t-k}^t), \text{ for any } t,$$

i.e., (S_k) .

Q.E.D.

PROOF OF THEOREM 1: It follows from Lemma 1 that (G_k) is equivalent to $\{(G_r), r=1,2,\dots\}$, i.e., to:

$$X_{t+1}^{t+r} \perp Y_{-\infty}^t \mid X_{-\infty}^t, \text{ for any } t, \text{ for any } r,$$

i.e., to (C).

Similarly, from Lemma 2 it follows that (S_h) is equivalent to $\{(S_r); r=1,2,\dots\}$. It now suffices to show that $\{(S_r); r=1,2,\dots\}$ is equivalent to (C).

From the definition of (C) and the FPCI, it is clear that (C) implies (S_r) for any r . To see the converse, it suffices to choose for every r , $Y_{t-r} = d$. Then

$$X_{t+1}^{\infty} \perp Y_{t-r+1}^t \mid X_{-\infty}^t, \text{ for any } t, \text{ for any } r,$$

which implies

$$X_{t+1}^{\infty} \perp Y_{-\infty}^t \mid X_{-\infty}^t, \text{ for any } t,$$

i.e., (C).

Q.E.D.

PROOF OF LEMMA 3: This directly follows from the FPCI by putting $A = X_{t+1}^{\infty}$, $B = Y_{-\infty}^t$, $C = X_{-\infty}^{t-m}$, and $D = X_{t-m+1}^t$.

Q.E.D.

PROOF OF THEOREM 2: By putting $A = X_{t+1}^T$, $B = Y_1^t$, $C = X_1^{t-m}$, and $D = X_{t-m+1}^t$, it follows from the FPCI that (R_m^T) is equivalent to:

$$X_{t+1}^T \perp X_1^{t-m} \mid X_{t-m+1}^t, t=m+1,\dots,T-1, \quad (\text{A.5})$$

and

$$X_{t+1}^T \perp Y_1^t \mid X_1^t, t=m,\dots,T-1. \quad (\text{A.6})$$

Since (A.5) is just (M_m^T) , it now suffices to show that (A.6) is equivalent to (ii) and (iii).

It is clear that (A.6) implies (ii) and (iii). To see the converse, we first note that (ii) is (A.6) written for $t=m$. The proof now proceeds by induction on t . Suppose that (A.6) holds for $t-1$ where $m \leq t-1 \leq T-2$, i.e.,

$$X_t^T \perp Y_1^{t-1} \mid X_1^{t-1}.$$

This implies

$$X_{t+1}^T \perp Y_1^{t-1} \mid X_1^t.$$

Since (s_t^T) holds for $m \leq t \leq T-1$, it follows from the FPCI that:

$$X_{t+1}^T \perp Y_1^t \mid X_1^t.$$

Q.E.D.

II. The proofs of the results of Section 4 implicitly use the theory of log-linear probability models (see e.g., S. J. Haberman (1974), Q. H. Vuong (1982)).

PROOF OF LEMMA 4: To establish (4.2), one can first show that the joint probability model for the qualitative variables X_1, \dots, X_T associated with the restrictions (M_m^T) is a hierarchical log-linear probability (LLP) model generated by the configurations $(X_1^{m+1}), (X_2^{m+2}), \dots, (X_{T-m}^T)$. Lemma 4 then follows from the fact that this hierarchical LLP model is decomposable (see S. J. Haberman (1974, Definition 5.4, p. 166)) so that one can apply successively Haberman's result on closed-form M.L. estimates (S. J. Haberman (1974), Theorem 5.1, p. 175)).

Alternatively, a direct proof consists in noting that (M_m^T) is equivalent to:

$$\{X_{t+m+1} \perp X_1^t \mid X_{t+1}^{t+m}; \text{ for any } t=1, \dots, T-m-1\} \quad (\text{A.7})$$

(This follows by successive application of the FPCI.) It now suffices to consider the recursive system of LLP models associated with the decomposition:

$$\Pr(X_1^T) = \Pr(X_1^{m+1}) \prod_{t=1}^{T-m-1} \Pr(X_{t+m+1} \mid X_1^{t+m}) \quad (\text{A.8})$$

Since there are no restrictions on $\Pr(X_1^{m+1})$, the joint probability model for X_1^{m+1} is saturated. Hence the M.L. estimate of $p(i_1^{m+1})$ is $n(i_1^{m+1})/n$. For every $t=1, \dots, T-m-1$, the only restriction is that X_1^t be excluded from the conditional model for X_{t+m+1} given X_1^{t+m} . It follows that the M.L. estimate of $\Pr(X_{t+m+1} \mid X_1^{t+m})$ can be obtained by considering the conditional saturated model for X_{t+m+1} given X_{t+1}^{t+m} . Hence the M.L. estimate of $p(i_{t+m+1} \mid i_{t+1}^{t+m})$ is $n(i_{t+m+1}^{t+m+1})/n(i_{t+1}^{t+m})$.

Since the M.L. estimate of $\Pr(X_1^T)$ subject to the restrictions (M_m^T) is simply the product of the above M.L. estimates, Equation (4.2) follows.

Q.E.D.

PROOF OF THEOREM 3: Since the M.L. estimate of $\Pr(X_1^T)$ under no restriction is simply $n(i_1^T)/n$, it is easy to see that LR_0^m as defined by Equation (4.3) is the LR statistic for testing (M_m^T) against the hypothesis of no restriction.

To derive the number of degrees of freedom ddf_0^m of that statistic, it suffices to count the number of independent restrictions that are imposed by (M_m^T) on $\Pr(X_1^T)$. One can show that the dimension of the model space of the LLP model for X_1^T associated with the restrictions (M_m^T) is equal to the term in brackets in (4.4) so that ddf_0^m is indeed given by (4.4). Alternatively, one can use the recursive decomposition (A.7). For every $t=1, \dots, T-m-1$,

$$\Pr(X_{t+m+1} \mid X_1^t, X_{t+1}^{t+m}) = \Pr(X_{t+m+1} \mid X_{t+1}^{t+m}), \text{ where } X_k \text{ has } I_k$$

categories. Since there are $(I_{t+m+1} - 1) \prod_{k=1}^{t+m} I_k$ independent conditional probabilities $p(i_{t+m+1} \mid i_1^t, i_{t+1}^{t+m})$ and $(I_{t+m+1} - 1) \prod_{k=t+1}^{t+m} I_k$ independent conditional probabilities $p(i_{t+m+1} \mid i_{t+1}^{t+m})$, the number of restrictions imposed by (M_m^T) is

$$\text{ddf}_0^m = \sum_{t=1}^{T-m-1} \left[(I_{t+m+1} - 1) \left(\prod_{k=1}^{t+m} I_k - \prod_{k=t+1}^{t+m} I_k \right) \right]$$

which, after simplification, gives (4.4).

Q.E.D.

PROOF OF COROLLARY 1: Obvious.

PROOF OF LEMMA 5: The only restriction on $\Pr(Y_1^m \mid X_1^m, X_{m+1}^T)$ is that $\Pr(Y_1^m \mid X_1^T) = \Pr(Y_1^m \mid X_1^m)$. It follows that the M.L. estimate of $p(j_1^m \mid i_1^T)$ is given by (4.11).

For every $t = m+1, \dots, T-1$, the only restriction on $\Pr(y_t \mid X_1^t, X_{t+1}^T, Y_1^{t-1})$ is that $\Pr(y_t \mid X_1^t, X_{t+1}^T, Y_1^{t-1}) = \Pr(y_t \mid X_1^t, Y_1^{t-1})$. It follows that the M.L. estimate of $p(j_t \mid i_1^t, j_1^{t-1})$ is given by (4.12).

Q.E.D.

PROOF OF THEOREM 4: From Theorem 2 and the recursive decomposition (4.7), it follows that the M.L. estimate of $\Pr(X_1^T, Y_1^{T-1})$ under the restrictions (R_m^T) is given by the right-hand side of (4.7) where the

joint and conditional probabilities are replaced respectively by their estimated joint and conditional probabilities obtained in Lemmas 4 and 5. Since the M.L. estimate of $\Pr(X_1^T, Y_1^{T-1})$ under no restrictions is given by:

$$\hat{p}(i_1^T, j_1^{T-1}) = \frac{n(i_1^T, j_1^{T-1})}{n} = \frac{n(i_1^T)}{n} \cdot \frac{n(i_1^T, j_1^m)}{n(i_1^T)} \cdot \prod_{t=m+1}^{T-1} \frac{n(i_1^T, j_1^t)}{n(i_1^T, j_1^{t-1})},$$

it follows from Equation (4.8)-(4.10) that the log-likelihood ratio statistic for testing (R_m^T) against the hypothesis of no restrictions is given by (4.13-4.15).

To compute the number of degrees of freedom of this statistic, it now suffices to count the number of restrictions imposed by (R_m^T) . From Theorem 3, we know that (M_m^T) imposes ddf_0^m restrictions on $\Pr(X_1^T)$. In addition, (c_m^T) requires that $\Pr(Y_1^m \mid X_1^m, X_{m+1}^T) = \Pr(Y_1^m \mid X_1^m)$ which introduces ddf_m^m restrictions where ddf_m^m is given by (4.17). Finally, for every $t = m+1, \dots, T-1$, (s_t^T) requires that $\Pr(y_t \mid X_1^t, X_{t+1}^T, Y_1^{t-1}) = \Pr(y_t \mid X_1^t, Y_1^{t-1})$ which introduces ddf_t^m restrictions where ddf_t^m is given by (4.18). From Theorem 2, it follows that the total number of restrictions imposed by (R_m^T) is given by (4.16).

Q.E.D.

PROOF OF COROLLARY 2: Obvious.

FOOTNOTES

- * This research was done while the third author was visiting the Universite des Sciences Sociales de Toulouse. Support from DGRST 81-E-1303 is gratefully acknowledged. We are indebted to Jeff Dubin, Dave Grether, Jerry Kramer, Doug Rivers, Howard Rosenthal for helpful criticism, and to Ken McCue for computational assistance.
1. To be rigorous, $A \perp B \mid C$ actually means that the σ -fields A and B are conditionally independent given the σ -field C (see e.g. M. Loeve (1954), A. Monfort (1980), for a definition of independence on σ -fields). Then X_r^S is the σ -field generated by the random variables x_t , $r \leq t \leq s$.
 2. A similar definition appears in R. Kohn (1981, p. 130) for the linear prediction case. See also Definition 1.b of J. P. Florens and M. Mouchart (1982, p. 585) for the general case.
 3. As a matter of fact, these authors do not use the linear predictor version of (S) but Sims' initial definition requiring that the linear predictor of y_t based on $X_{-\infty}^{\infty}$ be identical to the linear predictor of y_t based on $X_{-\infty}^t$ only. G. Chamberlain (1982, p. 578) obtains Sims equivalence result as a corollary of his general result.
 4. Using a general result, J. P. Florens and M. Mouchart (1982) show that (G) is equivalent to (C). This equivalence is here obtained

as a consequence of Lemma 1 of which the proof is quite simple.

5. Note, however, that (C) is not equivalent to (S_k^*) where (S_k^*) is $X_{t+1}^{\infty} \perp y_{t-r} \mid (X_{-\infty}^t, Y_{t-k})$ for any $0 \leq r \leq k-1$, any $Y_{t-k} \subseteq Y_{-\infty}^{t-k}$, and any t . This can be seen by noting that (S_k^*) is not equivalent to (S) as the following example shows. This example also appears in G. Chamberlain (1982, p. 573). Let y_1, y_2 be independent Bernoulli random variables with $\Pr(y_t = 1) = 1/2$ for $t = 1, 2$. Let $x_3 = y_1 y_2$, and let all the other variables be identically null. Then, x_3 is independent of y_1 , and x_3 is independent of y_2 so that (S_k^*) holds for any $k \geq 2$. On the other $x_3 \not\perp y_2 \mid y_1$ so that (S) does not hold. Note also that the non-equivalence between (S_k^*) and (S) implies from Lemma 2 that (S_k^*) and (S_k) are not equivalent.
6. This can readily be shown for the case in which the variables are all dichotomous. One can then use the theory of log-linear probability models (see e.g., M. Nerlove and S. J. Press (1976), Q. H. Vuong (1982)) to show that the joint probability model for the observed two dichotomous variables is saturated. It is worth noting that the possible non-identification of (C) does not necessarily follow from the well-known result that two observed variables, conditionally independent given an unobserved variable, may actually appear dependent.
7. In particular, the stationarity assumption allows one to integrate out the unobserved part of X in order to derive the

restrictions that are imposed by (C) on the observed random variables of the sample. See also J. J. Heckman (1981)'s discussion of the problem of initial conditions and its consequences on the estimation of a discrete time-discrete data stochastic process.

8. If X is a stochastic process of mutually independent random variables, then X is a Markov process of order zero. (It can in fact be shown that the converse is true if and only if any x is independent of the infinite past of X .) One may also assume that m is a non-negative real number. Then, Lemma 3 still holds. On the other hand, Theorem 2 and the results of Section 3 no longer hold when m is not an integer. This is so because the X and Y processes are observed discretely. Hence if m is not an integer, the discretely observed process X is not an AR but an ARMA process (see e.g., M. S. Phadke and S. M. Wu (1974)).
9. The equivalence between (M_m) and AR(m) is analogous to the equivalence result between (C) and (G).
10. One may think that (R_m^T) is not the set of all possible restrictions implied by (R_m) . This may be true only if one is willing to introduce additional assumptions on the X and Y processes. For instance, when $m = 2$, one may think that the restriction $X_2^T \perp Y_1 \mid (x_0, x_1)$ must be considered since X_2^T and Y_1 are both observed, even though x_0 is not. From the same argument as the one given in footnote 7, it however follows that such a

restriction does not imply any restrictions on $\Pr(X_1^T, Y_1^T)$ unless some further assumptions are introduced.

1. The proof of Theorem 2 shows that (ii) and (iii) are also equivalent to the set of restrictions (C_m^T) where $(C_m^T) = \{X_{t+1}^T \perp Y_1^T \mid X_1^t \text{ for any } t = m, \dots, T-1\}$. This set is simply the set of restrictions imposed by (C) on the observed variables, as if the X -process was starting at $t = 1$.
2. Note that I_t and J_t may depend on t . The only assumption is that they are finite. This is satisfied if the set of values for which x_t and y_t have non-zero probabilities is finite.
3. For theoretical references on log-linear probability models, see e.g., Y. M. Bishop, S. E. Fienberg, and P. W. Holland (1975), L. A. Goodman (1978), and S. J. Haberman (1974).
4. If $n(i_{t+1}^{t+m}) = 0$ for some t , then $n(i_t^{t+m}) = 0$. Lemma 5 also says that if we restrict ourselves to strictly positive probabilities, then the M. L. estimates of $p(i_1^T)$ under the restrictions (M_m^T) exist if and only if there are no empty cells in any of the $T - m - 1$ marginal contingency tables $\{X_2^{m+1}, \dots, X_{T-m}^{T-1}\}$. It is well known that this latter condition is necessary. That the condition is also sufficient follows from the particular log-linear probability model representing (M_m^T) . (For further details on the existence of M. L. estimates in joint log-linear probability models, see S. J. Haberman (1974), J. P. Link

(1983).) The convention $0 \div 0 = 0$ essentially allows the $p(i_1^T)$'s to be null, and correspond to the notion of extended M. L. estimates (S. J. Haberman (1974)).

15. T. W. Anderson and L. A. Goodman (1957) derives the Pearson chi-square statistic and LR statistic for testing the same hypotheses, but under the additional assumptions that $I_t = I$ (say) for any t , and X is a stationary process. Their treatment of the initial conditions is also somewhat different from the one given here.
16. This crucially depends on the fact that the set of joint distribution $\Pr(X_1^T, Y_1^{T-1})$ that satisfy (R_m^T) is equal to the set of distributions $\Pr(X_1^T, Y_1^{T-1})$ such that $\Pr(X_1^T)$ satisfies (M_m^T) , $\Pr(Y_1^m | X_1^T)$ satisfies (c_m^T) and $\Pr(y_t | X_1^T, Y_1^{t-1})$ satisfies (s_t^T) for every $t = m+1, \dots, T-1$. This is precisely the meaning of Theorem 2.
17. Actually, the survey has also been conducted since November 78, but with a different periodicity. For a more detailed discussion of the data, see e.g. M. B. Bouissou, J. J. Laffont and Q. H. Vuong (1983).
18. The implicit assumption is that good markets are isolated from each other so that one can simultaneously observe an excess demand on one market and an excess supply on another market. For a motivation of such an assumption, see e.g. J. Muellbauer

(1978).

19. There may be some problems with the interpretation to give to these answers. Previous work (M. B. Bouissou, J. J. Laffont and Q. H. Vuong (1983)) has shown that this interpretation is satisfactory. Moreover, alternative and more complex ways of using the answers do not change the qualitative features of the following results.
20. Though in principle, the answer to the price variation question should be treated as a continuous variable, the certainty of reported answers are questionable since individuals tend to round off their answers. As in earlier work (see e.g. B. Ottenwaelter and Q. H. Vuong (1982)) the categorization used is: if x denotes the reported percentage change, then " $x \geq 5$ ", " $0 < x \leq 5$ ", and " $x \leq 0$ " corresponds respectively to IP=1, IP=2, and IP=3. The category IP=2 then corresponds to a price stability in real terms after having taken into account the average inflation rate over the years 74-78.
21. This was due to the fact that we were unable to accept a Markov of order 1 for any series of 3 successive surveys when considering all the firms. Given that the average number of firms answering successive surveys drops from about 1000 to about 600 when going from 3 successive surveys to 4 successive surveys (the minimum number of periods required to test a Markov of order 2), and given that the dimension of the relevant contingency

table for $T=4$ is $3^4 \times 2^3$, i.e. 648, our non-causality tests which are based on large samples then become unjustified.

22. These results were obtained from the FORTRAN program CAUSE9 which is available from the authors. This program can accept as an input a raw file that contains missing observations, and in addition can select the desired subsample. The program is written so that the computer storage required is a multiple of the minimum of the number of cases and the dimension of the analyzed contingency table. Each of the presented analyses took about 30 seconds of CPU time.
23. As a matter of fact, our tests of E-non-causality entail here a loss of information since they do not use the available information on ID_{t-2} .

REFERENCES

- Anderson, T. W., and L. Goodman: "Statistical Inference about Markov Chains," Annals of Mathematical Statistics, 28 (1957), 89-110.
- Barro, R., and H. Grossman: Money, Employment and Inflation. Cambridge: Cambridge University Press, 1976.
- Benassy, J. P.: The Economics of Disequilibrium. Academic Press, 1982.
- Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland: Discrete Multivariate Analysis: Theory and Practice. Cambridge, Massachusetts: MIT Press, 1975.
- Bouissou, M. B., J.J. Laffont, and Q. H. Vuong: "Econometrie du Desequilibre sur Donnees Microeconomiques," Discussion Paper No 8308, GREMAQ, Universite des Sciences Sociales de Toulouse, 1983.
- Chamberlain, G.: "The General Equivalence of Granger and Sims Causality," Econometrica, 50 (1982), 569-581.
- Florens, J. P., and M. Mouchart: "A Note on Noncausality," Econometrica, 50 (1982), 583-606.
- Goodman, L. A.: Analyzing Qualitative/Categorical Data. Cambridge, Massachusetts: Abt Books, 1978.

- Granger, C. W. J.: "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," Econometrica, 37 (1969), 424-438.
- _____ : "Testing for Causality--A Personal Viewpoint," Journal of Economic Dynamics and Control, 2 (1980), 329-352.
- Haberman, S. J.: The Analysis of Frequency Data. Chicago: University of Chicago Press, 1974.
- Heckman, J. J.: "Statistical Models for Discrete Panel Data," in Structural Analysis of Discrete Data with Econometric Applications, ed. by C. Manski and D. McFadden. Cambridge, Massachusetts: MIT Press, 1981.
- Hosoya, Y.: "On the Granger Condition for Non-Causality," Econometrica, 45 (1977), 1735-1736.
- Kohn, R.: "A Characterization of Granger-Sims Exogeneity," Economic Letters, 8 (1981), 129-133.
- Laffont, J. J.: "Fix-Price Models: A Survey of Recent Empirical Work," Discussion Paper No 8305, GREMAQ, Universite des Sciences Sociales de Toulouse, 1983.
- Link, J. P.: "The Existence of the Maximum-Likelihood Estimate in Log-Linear Probability Models," Ph.D. Thesis, Department of Economics, Northwestern University,

- forthcoming, 1983.
- Loève, M.: Probability Theory. New York: D. Van Nostrand Company, 1954.
- Malinvaud, E.: The Theory of Unemployment Reconsidered. New York: John Wiley and Sons, 1977.
- Monfort, A.: Cours de Probabilite's. Paris: Economica, 1980.
- Muellbauer, J.: "Macrotheory vs Macroeconometrics: The Treatment of Disequilibrium in Macro Models," Birbeck college, 1978.
- Nerlove, M., and S. J. Press: "Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data," Discussion Paper No. 1, Center for Statistics and Probability, Northwestern University, 1976.
- Ottenwaelter, B., and Q. H. Vuong: "An Empirical Analysis of Backlog, Inventory, Production, and Price Adjustments: An Application of Recursive Systems of Log-Linear Models," Working Paper 442, California Institute of Technology, 1982.
- Phadke, M. S., and S. M. Wu: "Modeling of Continuous Stochastic Processes from Discrete Observations with Application to Sunspots Data," Journal of the American Statistical Association, 69 (1974), 325-329.

Simon, H. A.: "Causal Ordering and Identifiability," in Studies in Econometric Methods, ed. by W. C. Hood and T. C. Koopmans. Cowles Commission Monograph 14, New York, 1953.

Sims, C. A.: "Money, Income, and Causality," The American Economic Review, 62 (1972), 540-552.

Strotz, R. H., and H. Wold,: "Recursive versus Non-Recursive Systems: An Attempt at Synthesis," Econometrica, 28 (1960), 417-427.

Vuong, Q. H.: "Conditional Log-Linear Probability Models: A Theoretical Development with an Empirical Application," Ph.D. Thesis, Department of Economics, Northwestern University, 1982.