ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

TESTS OF NONCAUSALITY UNDER MARKOV
ASSUMPTIONS FOR QUALITATIVE PANEL DATA

By M. B. Bouissou, J. J. Laffont, and Q. H. Vuong[1]

For many years, social scientists have been interested in obtaining testable definitions of causality (Granger [12], Sims [28]). Recent works include those of Chamberlain [7] and Florens and Mouchart [8]. The present paper first clarifies the results of these latter papers by considering a unifying definition of noncausality. Then, log-likelihood ratio (LR) tests for noncausality 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 test statistics have closed forms. These tests thus provide a readily applicable procedure for testing noncausality 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.

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 Simon [27], Strotz and Wold [29]. 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 Granger [12] and Sims [28]. Recently, Chamberlain [7] and Florens and Mouchart [8] extended these latter definitions to possibly nonstationary nongaussian processes. The present paper first clarifies the results of these two recent papers, and second, derives some tests for noncausality 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 noncausality 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 Chamberlain's general equivalence result.

The essential difficulty in testing for noncausality is that the restrictions imposed by the noncausality 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, 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 are 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 our theoretical results.

2. SOME GENERAL RESULTS ON NONCAUSALITY

Let X and Y be two possibly nonstationary stochastic scalar or vector processes. In what follows, X and Y are discrete time processes, i.e., $\{(x_t, y_t): t \text{ in } Z\}$ where $Z = \{\ldots, -1, 0, 1, \ldots\}$. 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 noncausality 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 $A \perp B \mid C$.[2]

The definitions of noncausality that we consider are those of Granger [12] and of Sims [28] (as modified by Chamberlain [7]).

DEFINITION 1 (Granger Noncausality): The stochastic process Y does not *Granger cause* the stochastic process X if

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

---

[2] 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., Loeve [20], Monfort [22], for a definition of independence on σ-fields). Then X_r^s is the σ-field generated by the random variables x_t, $r \leq t \leq s$.

DEFINITION 2 (Sims–Chamberlain Strict Exogeneity): The stochastic process X is *strictly exogenous* to the stochastic process Y if

(S) $X^{\infty}_{t+1} \perp y_t \mid (X^t_{-\infty}, Y_{t-1})$

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

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 noncausality is sufficiently strong. This suggests the following definition of noncausality, which has also been considered by Florens and Mouchart [8].

DEFINITION 3 (Global Noncausality): The stochastic process Y does not *cause* the stochastic process X if

(C) $X^{\infty}_{t+1} \perp Y^t_{-\infty} \mid X^t_{-\infty}$,   for any t.

Global noncausality 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.[3]

The previous definitions apply to completely general discrete-time processes since the X and Y processes need not satisfy any particular assumptions.[4] 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., Sims [28] for covariance stationary processes with autoregressive representation and no linearly deterministic component, and Hosoya [16] for more general processes.[5]) Chamberlain [7], in addition to modifying Sims initial definition, establishes directly the equivalence between (G) and (S) in the general case.

The remainder of this section provides an indirect but, we think, clarifying proof of Chamberlain's general equivalence result [7, Theorem 4]. Moreover, our derivation will produce a general result of which Chamberlain's equivalence is a special case. We need some additional definitions and some lemmas. Let $k \geqslant 1$.

---

[3] A similar definition appears in Kohn [17, p. 130] for the linear prediction case.

[4] These definitions can be extended to continuous-time processes as follows. Let $X^{t-}_{-\infty}$ 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 $Y^{t-}_{-\infty}$ is similarly defined. The previous definitions apply to continuous-time processes provided "$t-1$" and "$t+1$" are respectively replaced by "$t-$" and "$t+$".

[5] 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 $X^t_{-\infty}$ only. Chamberlain [7, p. 578] obtains Sims equivalence result as a corollary of his general result.

DEFINITION 4 (Granger Noncausality of order k): The stochastic process Y does not *Granger cause at the order k* the stochastic process X if:

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

Granger noncausality 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 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 noncausality, i.e., (G_1), is equivalent to any (G_k).

Granger noncausality 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) $X_{t+1}^{\infty} \perp Y_{t-k+1}^t \,|\, (X_{-\infty}^t, Y_{t-k})$

for any subset Y_{t-k} 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).

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). 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.

There exist equivalent versions of (C) other than (G) and (S). For instance, one may consider the following apparently weaker forms of noncausality of Y on X.

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

(G_k^*) $x_{t+r} \perp Y_{-\infty}^t \,|\, X_{-\infty}^t$,   for any $1 \leq r \leq k$, 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) which implies (C).[6] It is noteworthy that this latter result implies that (G_k) and (G_k^*) are actually equivalent.

3. NONCAUSALITY UNDER MARKOV ASSUMPTIONS

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

The essential difficulty in testing for noncausality is that noncausality of Y on X involves a conditioning set with an infinite number of random variables, namely $X_{-\infty}^t$. Since in general one does not observe the whole part of X, noncausality of Y on X may not be statistically identified. To circumvent the problem of conditioning on sets with an infinite number of variables, one may simply assume that the X process starts at $t = 1$ (the first period of the sample), or 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. One may instead assume that the X and Y processes are jointly stationary, as it is usually done in econometric work.[7] Then, one must in general restrict the class of probability distributions to be considered, i.e., one specifies the probability model generating the stationary processes X and Y. It follows that the inference that one can make about noncausality is conditional upon the truthfulness of the additional assumptions put forward to identify (C).

Since the question of whether any statement can be made about noncausality based just on statistical data is important, as Granger [13] 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.[8] In particular, the X and Y processes need not be stationary. Moreover, the Y

---

[6] 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 $0 \le r \le k-1$, any $Y_{t-k} \subset Y_{-\infty}^{t-k}$, 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 Chamberlain [7, p. 573]. Let y_1, y_2 be independent Bernoulli random variables with $\Pr(y_t = 1) = \frac{1}{2}$ and $\Pr(y_t = -1) = \frac{1}{2}$ 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 $k \ge 2$. On the other hand $x_3 \not\perp y_2 \mid y_1$ so that (S) does not hold. Note also that the nonequivalence between (S_k^*) and (S) implies from Lemma 2 that (S_k^*) and (S_k) are not equivalent.

[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.

[8] Florens and Mouchart [9, Section 4] also considered Markov assumptions. Note, however, that Lemma 3 below is stronger than Florens and Mouchart Theorem 4.2 since our condition (R_m) clearly implies their conditions (FG) and (FS). Moreover, (FG) and (FS) contain common restrictions, while our characterization given in Theorem 2 does not.

process need not be Markov. This is simply because we are testing for the noncausality 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 noncausality of Y on X and the Markov requirement on X.

Let m be an integer possibly equal to zero.[9] 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) $X^\infty_{t+1} \perp X^{t-m}_{-\infty} \mid X^t_{t-m+1}$,   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.

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 $m \geqslant 0$, (C) and (M_m) both hold if and only if (R_m) holds, where*

(R_m) $X^\infty_{t+1} \perp (X^{t-m}_{-\infty}, Y^t_{-\infty}) \mid X^t_{t-m+1}$,   *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 a 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^T_1, Y^T_1).

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 $T \geqslant m + 2$. Indeed, if this were not the case, we

[9] If X is a stochastic process of mutually independent random variables, then X is a Markov process of order zero. One may also assume that m is a nonnegative real number. However, Theorem 2 and the results of Section 3 no longer hold since if m is not an integer, the discretely observed process X is no longer an AR process (see, e.g., Phadke and Wu [26]).

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:

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

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 noncausality of Y on X and the Markov assumption on X. It can readily be seen that these restrictions are:

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

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.

The next theorem presents the basic result that underlies the tests for noncausality 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 $m \geqslant 0$, (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* $t = m+1, \ldots, T-1$: (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 $t \geqslant m+1$, each condition (s_t^T) is Sims condition written at t only, as if the X and Y processes were both starting at $t = 1$.[10]

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}^T conditional upon all the other observed x's. Condition

---

[10] The proof of Theorem 2 shows that (ii) and (iii) are also equivalent to the set of restrictions (C_m^T) where $(C_m^T) = \{X_{t+1}^T \perp Y_1^t \mid X_1^t$ for any $t = m, \ldots, T-1\}$. This set is simply the set of restrictions imposed by (C) on the observed variables, as if the X process was starting at $t = 1$.

(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 nondependence 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 $t = m+1, \ldots, T-1$.

4. TESTS OF NONCAUSALITY 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 allow 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 distributions $\Pr(X_1^T, Y_1^T)$ 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 nonparametric tests based on goodness-of-fit statistics such as log-likelihood ratio (LR) statistics (see, e.g., Goodman [11], Haberman [14]), 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.[11]

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 $t = 1, \ldots, T$, 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 noncausality 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 $i_1^T = (i_1, \ldots, i_T)$ and $j_1^{T-1} = (j_1, \ldots, j_{T-1})$, 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, $p(i_r^s, j_t^u)$ denotes the probability that X_r^s and Y_t^u are respectively equal to i_r^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 statistic. 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 $\{n(i_r^s, j_t^u),$ for any

---

[11] Though there exist other goodness-of-fit statistics such as the Pearson chi-square statistics, we shall consider LR statistics only. This is so because LR tests are known to be most powerful in the sense of Hoeffding's asymptotic efficiency (Hoeffding [15]) and Bahadur's exact slopes (Bahadur [2]).

[12] Note that I_t and J_t may depend on t. The only assumption is that they are finite.

$(i_r^s, j_t^u)\}$ 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 noncausality 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:

$$(4.1) \qquad \log L_0 = \sum_{i_1^T} n(i_1^T) \log p(i_1^T).$$

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 (ML) estimates of the probabilities $p(i_1^T)$ under the restrictions (M_m^T). The import of the result is that the ML estimates have a closed form so that they can readily be computed.[13]

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

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

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

$$(4.3) \qquad LR_0^m = 2 \sum_{i_1^T} n(i_1^T) \log \frac{n(i_1^T)}{n \hat{p}^m(i_1^T)}.$$

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_0^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*

[13] The lemma follows from 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. For theoretical references on log-linear probability models, see e.g., Bishop, Fienberg, and Holland [5], Goodman [11], and Haberman [14].

[14] 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 ML 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 $\{X_2^{m+1}\}, \ldots, \{X_{T-m}^{T-1}\}$. 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 ML estimates in log-linear probability models, see Haberman [14], Link [17].) The convention $0 \div 0 = 0$ allows the $p(i_1^T)$'s to be null and correspond to the notion of extended ML estimates (Haberman [14]).

null hypothesis, this statistic follows a chi-square distribution with number of degrees of freedom:

$$(4.4) \qquad ddf_0^m = \left(\prod_{t=1}^{T} I_t \right) - \left[\sum_{t=1}^{T-m} \left(\prod_{k=t}^{t+m} I_k \right) - \sum_{t=1}^{T-m-1} \left(\prod_{k=t+1}^{t+m} I_k \right) \right].$$

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

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

where $\hat{p}^r(i_1^T)$ is the ML 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*

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

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

We now turn to the testing of the noncausality 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

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

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

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

$$\log L = \sum_{(i_1^T, j_1^{T-1})} n(i_1^T, j_1^{T-1}) \log p(i_1^T, j_1^{T-1})$$

$$(4.8) \qquad = \log L_0 + \log L_m + \sum_{t=m+1}^{T-1} \log L_t$$

---

[15] Anderson and Goodman [1] derive 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.

where $\log L_0$ is given by (4.1), and

$$\log L_m = \sum_{(i_1^T, j_1^m)} n(i_1^T, j_1^m) \log p(j_1^m \mid i_1^T),$$

$$\log L_t = \sum_{(i_1^T, j_1^t)} n(i_1^T, j_1^t) \log p(j_t \mid i_1^T, j_1^{t-1}),$$

for any $t = m + 1, \ldots, T$. Hence the log-likelihood function $\log L$ is simply the sum of the marginal and conditional log-likelihood functions associated with the probability models composing the recursive system. It follows that the ML 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 subject to the appropriate constraints.[16]

The next lemma gives the ML 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 ML 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 ML estimate of $p(j_1^m \mid i_1^T)$ under the restrictions (c_m^T) is*

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

and for any $t = m + 1, \ldots, T - 1$ and for any (i_1^T, j_1^T), the ML estimate of $p(j_t \mid i_1^T, j_1^{t-1})$ is

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

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

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

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

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

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

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

---

[16] This crucially depends on the fact that the set of joint distributions $\Pr(X_1^T, Y_1^{T-1})$ 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), $\Pr(Y_1^m \mid X_1^T)$ satisfies (c_m^T), and $\Pr(y_t \mid X_1^T, Y_1^{t-1})$ satisfies (s_t^T) for every $t = m + 1, \ldots, T - 1$. This is precisely the meaning of Theorem 2. In fact, this system is a recursive system of conditional log–linear probability models (see Vuong [30]).

THEOREM 4 (LR Test for Noncausality and Markov of Order m): *For any m such that $0 \le m \le T-2$, 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 with number of degrees of freedom:*

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

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

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

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

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

One may also want to test that Y does not cause X under the maintained hypothesis that X is Markov of order m. Let

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

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 Noncausality under Markov of Order m): *For any m such that $0 \le m \le T-2$, 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:*

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

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 $(c_m^T), (s_{m+1}^T), \ldots, (s_{T-1}^T)$ that are imposed by the noncausality 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 EXAMPLE

Since the initial theoretical work in disequilibrium economics of Barro and Grossman [3], Benassy [4], and Malinvaud [21], fix-price models have been

estimated frequently (see Laffont [18] for a survey of recent empirical work). The fix-price paradigm does not, however, imply that prices never change:

"... 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." (Malinvaud [21, p. 12].)

The purpose of this section is to illustrate our previous results by testing 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.

The data that we use have been collected by the Institut National de la Statistique et des Etudes Economiques (INSEE) from about 3,000 firms through periodic Business Survey Tests taken each year in March, June, and November, starting from June, 1974 to June, 1978.[17] As a matter of fact, the sample is much smaller due to missing observations. For instance, the average number of firms in all sectors answering three successive surveys drops to about 1,000.[18]

We shall be interested in the disequilibrium experienced by each firm on its good market.[19] 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, while if the firm answers NO we presume that there is excess demand.[20]

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

[17] Actually, the survey has also been conducted since October 1978, but with a different periodicity. For a more detailed discussion of the data, see Bouissou, Laffont, and Vuong [6]).

[18] This will often arise with survey data since individuals (firms in our case) will not answer all successive surveys. Moreover, for a given survey, these individuals will not in general answer all relevant questions. Hence, the larger the number T of periods, the smaller will be the sample of complete observations. On the other hand, the number of periods must be sufficiently large so as to be able to test that X is Markov of order k, specifically $T \geq k - 2$. It follows that, when k increases, the sample size will become small relative to the dimension of the analyzed contingency table which is $I^T \cdot J^{T-1}$. This conflicts with the large sample nature of our tests: As a rule-of-thumb, it is recommended that the sample size be at least four or five times the dimension of the contingency table.

A solution is then to consider methods for using efficiently the incomplete observations. Though this problem is important, it is beyond the scope of this paper. A possible approach, however, is to use recent results on maximum-likelihood estimation in contingency tables with randomly missing data (Fuchs [10]) where the EM algorithm is suggested to solve the modified normal equations.

[19] 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 Muellbauer [23].

[20] There may be some problems with the interpretation given to these answers. Previous work (Bouissou, Laffont, and Vuong [6]) has shown that this interpretation is satisfactory.

first category, is constructed so that it corresponds to an increase in real terms; the second category, to a stability; and the third category, to a decrease.[21]

The question 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. Using the sample of about 1,000 firms that answer a series of three successive surveys, we were, however, unable to accept a Markov of order 1 for the IP process.[22] This can be explained by the presence of firms such as those in the professional equipment sector for which the price adjustment process is expected to have much more memory than three months as can be seen from Tables I and II. We have then restricted our analysis to the consumption good sector.

Table I presents our results for the consumption sector when analyzing three successive surveys.[23] As a basis for comparison, Table II presents similar results for the whole industry with randomly selected subsamples of sizes equal to the sizes of the corresponding samples for the consumption sector.[24] 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.

---

[21] Though in principle, the answer to the price variation question should be treated as a continuous variable, the continuity of reported answers are questionable since individuals tend to round off their answers. As in earlier work (see, e.g., Ottenwaelter and Vuong [25]) the categorization used is: if x denotes the reported percentage change, then "$x \geqslant 5$", "$0 < x \leqslant 5$", and "$x \leqslant 0$" corresponds respectively to $IP = 1$, $IP = 2$, and $IP = 3$. The category $IP = 2$ then corresponds to a price stability in real terms having taken into account the average inflation rate over the years 1974–1978.

[22] We could not test a Markov of order 2 on the IP process. Indeed the average number of firms answering 4 successive surveys drops to about 600 where 4 is the minimum number of periods required to test a Markov of order 2. Then the sample size becomes about equal to the dimension of the contingency table for $T = 4$ which is $3^4 \times 2^3 = 648$ (see footnote 18).

[23] 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.

[24] Our reason for considering samples of the same size for the whole industry as for the consumption sector is to have a fair comparison. Indeed, it is well-known that LR tests are consistent (and in our case, even optimal; see footnote 10). Thus for fixed (asymptotic) probability of type I error, the probability of type II error goes to zero as the sample size increases. Hence our LR tests will reject the null hypothesis (M) or (C) even when the true model departs only slightly from the null hypothesis which is in general a simplification of the world. In other words, it is more likely to reject the null hypothesis with a large sample than with a small sample.

Instead of reducing the samples for the whole industry, an alternative method would be to choose critical levels depending on the sample size so that the probabilty of type I error goes to zero while the probability of type II error either goes to a constant between zero and one (Bahadur [2]) or to zero (Hoeffding [15]) for a fixed alternative. There are, however, no generally accepted ways of choosing how the sizes of the test should go to zero when the sample size increases.

TABLE I

LR Statistics with Upper-Tail Probabilities in Parentheses

Consumption Sector

Ending Periods $TE = 3$	Number of Cases	For Markov of Order 1 on X $DF = 12$	For Noncausality of Y on X assuming Markov of Order 1 on X $DF = 60$	For Noncausality of Y on X and for Markov of Order 1 on X $DF = 72$
75-03	413	12.5*	59.2*	71.7*
		(40.8)	(50.4)	(48.8)
75-06	397	16.5*	37.7*	54.1
		(17.1)	(98.9)	(94.2)
75-11	386	30.5	30.8*	61.3*
		(.237)	(99.9)	(81.1)
76-03	387	12.6*	60.2*	72.8*
		(39.8)	(46.7)	(45.0)
76-06	398	32.8	68.9*	101.7
		(.103)	(20.2)	(1.22)
76-11	384	52.1	72.2*	124.2
		(.000)	(13.4)	(0.13)
77-03	345	8.9*	68.9*	77.8*
		(71.2)	(20.2)	(30.0)
77-06	356	13.4*	59.3*	72.7*
		(33.9)	(50.0)	(45.2)
77-11	395	29.2	74.3*	103.5
		(.362)	(10.2)	(.887)
78-03	367	16.1*	65.1*	81.1*
		(18.5)	(30.4)	(21.5)
78-06	401	31.6	62.2*	93.7
		(.002)	(39.9)	(.044)

* Indicates that the null hypothesis cannot be rejected at the 10 per cent significance level.

For the consumption sector, we reject the hypothesis that the IP process is Markov of order 1 for five out of eleven periods at the 10 percent significance level, while for the whole industry that hypothesis is rejected for 10 out of 11 periods. Thus for these periods, nothing can be said about noncausality. On the other hand, for all the periods for which the first-order Markov assumption cannot be rejected at the 10 per cent significance level, the hypothesis that ID does not cause IP is always accepted. Thus our results seem to support the hypothesis that price changes from period to period are strictly exogenous to disequilibria appearing within periods.

6. CONCLUSION

In this paper, we have considered a unifying definition of noncausality which has been proved to be equivalent to Granger's definition of noncausality and to Chamberlain's revised version of Sims' strict exogeneity.

After having argued that noncausality of Y on X is by itself nonidentified in practice, we have introduced the additional assumption that X is Markov of

TABLE II

LR STATISTICS WITH UPPER-TAIL PROBABILITIES IN PARENTHESES

WHOLE INDUSTRY

Ending Periods $TE = 3$	Number of Cases	For Markov of Order 1 on X $DF = 12$	For Noncausality of Y on X Assuming Markov of Order 1 on X $DF = 60$	For Noncausality of Y on X and for Markov of Order 1 on X $DF = 72$
75-03	413	18.0*	43.5*	61.5*
		(11.7)	(94.6)	(80.6)
75-06	397	32.1	60.8*	92.9
		(.133)	(44.6)	(4.91)
75-11	386	39.6	44.2*	83.8*
		(.008)	(93.7)	(16.1)
76-03	387	26.2	41.2*	67.4*
		(1.01)	(96.9)	(63.2)
76-06	398	29.9	49.0*	78.9*
		(.287)	(84.4)	(26.9)
76-11	384	28.2	67.5*	95.7
		(.514)	(23.5)	(3.21)
77-03	345	37.7	65.7*	103.4
		(.017)	(28.6)	(.898)
77-06	356	30.8	53.7*	84.5*
		(.213)	(70.4)	(14.9)
77-11	395	42.9	38.0*	80.9*
		(.002)	(98.8)	(22.1)
78-03	367	20.2	49.3*	69.5*
		(6.30)	(83.6)	(56.0)
78-06	401	37.4	66.6*	104.0
		(.019)	(26.1)	(.812)

* Indicates that the null hypothesis cannot be rejected at the 10 per cent significance level.

some order. Then, using a recursive decomposition of all the restrictions that are imposed on panel data by the noncausality 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 noncausality 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 has been 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. With a sample size of about 400 firms, conditionally on the fact that price variations follow a Markov process of order one, the assumption that the disequilibrium indicator does not cause price variation cannot be rejected at the 10 per cent significance level, either in the consumption sector or in the whole industry. However, the data reject much more often the Markov assumption of order one in the whole industry than in the consumption sector.

Université des Sciences Sociales, Toulouse
and
California Institute of Technology

APPENDIX

1. The following fundamental property of conditional independence (FPCI) is used to prove the results of Section 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) $A \perp B \mid (C, D)$ and

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

(see, e.g., Florens and Mouchart [8, Theorem A.1, p. 588]).

PROOF OF LEMMA 1: (G_{k+1}) clearly implies (G_k). The proof of the converse is similar to the proof of Theorem 1 in Florens and Mouchart [8, p. 590].

PROOF OF LEMMA 2: (S_k) implies (S_{k+1}). Let Y_{t-k-1} be a subset of $Y^{t-k-1}_{-\infty}$. Since $Y_{t-k-1} \cup y_{t-k}$ is a subset of $Y^{t-k}_{-\infty}$, and since (S_k) holds at t, we have

$$X^\infty_{t+1} \perp Y^t_{t-k+1} \mid (X^t_{-\infty}, Y_{t-k-1}, y_{t-k}), \quad \text{for any } t$$

which implies from the FPCI:

(A.1) $X^\infty_{t+1} \perp y_t \mid (X^t_{-\infty}, Y_{t-k-1}, Y^{t-1}_{t-k}), \quad \text{for any } t.$

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

(A.2) $X^\infty_t \perp Y^{t-1}_{t-k} \mid (X^{t-1}_{-\infty}, Y_{t-k-1}), \quad \text{for any } t.$

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

$$X^\infty_{t+1} \perp Y^t_{t-k} \mid (X^t_{-\infty}, Y_{t-k-1}), \quad \text{for any } t,$$

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 $Y^{t-k-1}_{-\infty}$ so that from (S_{k+1}) we get

$$X^\infty_{t+1} \perp Y^t_{t-k} \mid (X^t_{-\infty}, Y_{t-k}), \quad \text{for any } t,$$

which implies (S_k), i.e.,

$$X^\infty_{t+1} \perp Y^t_{t-k+1} \mid (X^t_{-\infty}, Y_{t-k}), \quad \text{for any } t.$$

(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 $Y^{t-k-1}_{-\infty}$. From (S_{k+1}) it follows that

$$X^\infty_{t+1} \perp Y^t_{t-k} \mid (X^t_{-\infty}, Y_{t-k-1}), \quad \text{for any } t,$$

which implies

$$X^\infty_{t+1} \perp Y^t_{t-k+1} \mid (X^t_{-\infty}, Y_{t-k}), \quad \text{for any } t,$$

i.e., (S_k).   Q.E.D.

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

$$X^{t+r}_{t+1} \perp Y^t_{-\infty} \mid X^t_{-\infty}, \quad \text{for any } t, \text{ for any } r,$$

and hence to (C).

Similarly, from Lemma 2 it follows that (S_h) is equivalent to $\{(S_r); r = 1, 2, \ldots\}$. It now suffices to show that $\{(S_r); r = 1, 2, \ldots\}$ 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, $Y_{t-r} = \varnothing$. Then

$$X^{\infty}_{t+1} \perp Y^t_{t-r+1} \,|\, X^t_{-\infty}, \quad \text{for any } t, \text{ for any } r,$$

which implies

$$X^{\infty}_{t+1} \perp Y^t_{-\infty} \,|\, X^t_{-\infty}, \quad \text{for any } t,$$

i.e., (C).   Q.E.D.

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

PROOF OF THEOREM 2: By putting $A = X^T_{t+1}$, $B = Y^t_1$, $C = X^{t-m}_1$, and $D = X^t_{t-m+1}$, it follows from the FPCI that (R^T_m) is equivalent to:

(A.3) $X^T_{t+1} \perp X^{t-m}_1 \,|\, X^t_{t-m+1}$ $(t = m+1, \ldots, T-1)$,

and

(A.4) $X^T_{t+1} \perp Y^t_1 \,|\, X^t_1$ $(t = m, \ldots, T-1)$.

Since (A.3) is just (M^T_m), it now suffices to show that (A.4) is equivalent to (ii) and (iii).

It is clear that (A.4) implies (ii) and (iii). To see the converse, we first note that (ii) is (A.4) written for $t = m$. The proof now proceeds by induction on t. Suppose that (A.4) holds for $t-1$ where $m \leqslant t-1 \leqslant T-2$, i.e.,

$$X^T_t \perp Y^{t-1}_1 \,|\, X^{t-1}_1.$$

This implies

$$X^T_{t+1} \perp Y^{t-1}_1 \,|\, X^t_1.$$

Since (s^T_t) holds for $m \leqslant t \leqslant T-1$, it follows from the FPCI that

$$X^T_{t+1} \perp Y^t_1 \,|\, X^t_1.$$ Q.E.D.

II. The results of Section 4 can be proved using the theory of log–linear probability models (see, e.g., Haberman [14], Nerlove and Press [24], Vuong [30]). We shall, however, offer direct proofs.

PROOF OF LEMMA 4: Note that (M^T_m) is equivalent to

(A.5) $\{X_{t+m+1} \perp X^t_1 \,|\, X^{t+m}_{t+1}; \text{ for any } t = 1, \ldots, T-m-1\}$.

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

(A.6) $\displaystyle \Pr(X^T_1) = \Pr(X^{m+1}_1) \prod_{t=1}^{T-m-1} \Pr(X_{t+m+1} \,|\, X^{t+m}_1)$.

Since there are no restrictions on $\Pr(X^{m+1}_1)$, the joint probability model for X^{m+1}_1 is saturated. Hence the ML estimate of $p(i^{m+1}_1)$ is $n(i^{m+1}_1)/n$. For every $t = 1, \ldots, T-m-1$, the only restriction is that X^t_1 be excluded from the conditional model for X_{t+m+1} given X^{t+m}_1. It follows that the ML estimate of $p(i_{t+m+1} \,|\, i^{t+m}_{t+1})$ is $n(i^{t+m+1}_{t+1})/n(i^{t+m}_{t+1})$.

Since the ML estimate of $\Pr(X^T_1)$ subject to the restrictions (M^T_m) is simply the product of the above ML estimates, Equation (4.2) follows.   Q.E.D.

PROOF OF THEOREM 3: Since the ML estimate of $\Pr(X^T_1)$ under no restriction is simply $n(i^T_1)/n$, it is easy to see that LR^m_0 as defined by Equation (4.3) is the LR statistic for testing (M^T_m) 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 $\Pr(X_1^T)$. We now use the recursive decomposition (A.5). For every $t = 1, \ldots, T - m - 1$, $\Pr(X_{t+m+1} \mid X_1^t, X_{t+1}^{t+m}) = \Pr(X_{t+m+1} \mid X_{t+1}^{t+m})$, where X_k 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

$$ddf_0^m = \sum_{t=1}^{T-m-1} \left[(I_{t+m+1} - 1) \left(\prod_{k+1}^{t+m} I_k - \prod_{k=t+1}^{t+m} I_k \right) \right]$$

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 ML estimate of $p(j_1^m \mid i_1^T)$ is given by (4.11).

For every $t = m + 1, \ldots, T - 1$, 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 ML 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 ML 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 ML 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 $t = m + 1, \ldots, T - 1$, (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.

REFERENCES

[1] ANDERSON, T. W., AND L. GOODMAN: "Statistical Inference about Markov Chains," *Annals of Mathematical Statistics*, 28(1957), 89–110.

[2] BAHADUR, R. R.: "An Optimal Property of the Likelihood Ratio Statistics," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics*, Vol. 1. Berkeley: University of California Press, 1967, 13–26.

[3] BARRO, R., AND H. GROSSMAN: *Money, Employment and Inflation*. Cambridge: Cambridge University Press, 1976.

[4] BENASSY, J. P.: *The Economics of Disequilibrium*. New York: Academic Press, 1982.

[5] BISHOP, Y. M. M., S. E. FIENBERG, AND P. W. HOLLAND: *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Massachusetts: MIT Press, 1975.

[6] BOUISSOU, M. B., J. J. LAFFONT, AND Q. H. VUONG: "Econométrie du Déséquilibre sur Données Microéconomiques," *Annales de l'INSEE*, 55/56(1984), 109–151.

[7] CHAMBERLAIN, G.: "The General Equivalence of Granger and Sims Causality," *Econometrica*, 50(1982), 569–581.

[8] FLORENS, J. P., AND M. MOUCHART: "A Note on Noncausality," *Econometrica*, 50(1982), 583–606.

[9] ———: "A Linear Theory for Non-Causality," *Econometrica*, 53(1985), 157–175.

[10] FUCHS, C.: "Maximum Likelihood Estimation and Model Selection in Contingency Tables with Missing Data," *Journal of the American Statistical Association*, 77(1982), 270–278.

[11] GOODMAN, L. A.: *Analyzing Qualitative/Categorical Data.* Cambridge, Massachusetts: Abt Books, 1978.
[12] GRANGER, C. W. J.: "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica,* 37(1969), 424–438.
[13] ————: "Testing for Causality—A Personal Viewpoint," *Journal of Economic Dynamics and Control,* 2(1980), 329–352.
[14] HABERMAN, S. J.: *The Analysis of Frequency Data.* Chicago: University of Chicago Press, 1974.
[15] HOEFFDING, W.: "Asymptotic Optimal Tests for Multidimensional Distributions," *Annals of Mathematical Statistics,* 36(1965), 369–401.
[16] HOSOYA, Y.: "On the Granger Condition for Non-Causality," *Econometrica,* 45(1977), 1735–1736.
[17] KOHN, R.: "A Characterization of Granger-Sims Exogeneity," *Economic Letters,* 8(1981), 129–133.
[18] LAFFONT, J. J.: "Fix-Price Models: A Survey of Recent Empirical Work," Discussion Paper No. 8305, GREMAQ, Université des Sciences Sociales de Toulouse, 1983.
[19] LINK, J. P.: "The Existence of the Maximum-Likelihood Estimate in Log-Linear Probability Models," Ph.D. Thesis, Department of Economics, Northwestern University, 1984.
[20] LOÈVE, M.: *Probability Theory.* New York: D. Van Nostrand Company, 1954.
[21] MALINVAUD, E.: The Theory of Unemployment Reconsidered. New York: John Wiley and Sons, 1977.
[22] MONFORT, A.: *Cours de Probabilités.* Paris: Economica, 1980.
[23] MUELLBAUER, J.: "Macrotheory vs Macroeconometrics: The Treatment of Disequilibrium in Macro Models," Birbeck College, 1978.
[24] 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, North-western University, 1976.
[25] 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," *Journal of Business and Economic Statistics,* 2(1984), 224–234.
[26] 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.
[27] SIMON, H. A.: "Causal Ordering and Identifiability," in *Studies in Econometric Methods,* Cowles Commission Monograph 14, ed. by W. C. Hood and T. C. Koopmans. New York: 1953.
[28] SIMS, C. A.: "Money, Income, and Causality," *The American Economic Review,* 62(1972), 540–552.
[29] STROTZ, R. H., AND H. WOLD: "Recursive versus Non-Recursive Systems: An Attempt at Synthesis," *Econometrica,* 28(1960), 417–427.
[30] VUONG, Q. H.: "Conditional Log-Linear Probability Models: A Theoretical Development with an Empirical Application," Ph.D. Thesis, Department of Economics, Northwestern University, 1982.