J. R. Statist. Soc. B (1993)
55, *No.* 4, *pp.* 849–870

Multiregression Dynamic Models

By CATRIONA M. QUEEN† and JIM Q. SMITH

University of Nottingham, UK — *University of Warwick, Coventry, UK*

[Received July 1990. Final revision June 1992]

SUMMARY

Multiregression dynamic models are defined to preserve certain conditional independence structures over time across a multivariate time series. They are non-Gaussian and yet they can often be updated in closed form. The first two moments of their one-step-ahead forecast distribution can be easily calculated. Furthermore, they can be built to contain all the features of the univariate dynamic linear model and promise more efficient identification of causal structures in a time series than has been possible in the past.

Keywords: DYNAMIC LINEAR MODELS; GRANGER CAUSALITY; INFLUENCE DIAGRAMS; MULTIPROCESS MODELS; NON-LINEAR TIME SERIES; RECURSIVE SIMULTANEOUS EQUATION MODELS

1. INTRODUCTION

This paper introduces a class of multivariate state space time series models, called multiregression dynamic models (MDMs), defined on a multivariate time series $\{\mathbf{Y}_t\}_{t \geqslant 1}$. These models have been developed to forecast time series whose components are hypothesized to exhibit certain conditional independences in any one time frame and there is believed to be a causal driving mechanism within the system. To illustrate the type of time series which may have these properties, Section 5.1 considers the problem of forecasting the market share and relative market price of a particular brand in a product market. The causal drive through this particular system is clearly seen as the brand's relative price tends to dictate its market share.

MDMs define a class of non-Gaussian time series models which decompose the forecasting system into components whose conditional distributions are univariate Bayesian dynamic linear regression models (DLMs) (Harrison and Stevens, 1976). As in simultaneous equation models, the components of the vector $\mathbf{Y}_t$ are regressed on other contemporaneous components of that vector. Since the forecasting model is expressed in terms of univariate DLMs, facilities available for the modelling of univariate series, such as intervention, trends and seasonals, can be transferred directly on to these models with no appeal to stringent symmetry conditions being necessary (as in Harvey (1986) and Quintana and West (1987)). Although the modelled processes can be highly non-linear they are often easily updated in closed form. This makes the process especially interesting because the models are amenable to analytical investigation. Approximate or numerical methods, which are often required for non-Gaussian series (for example, Kitagawa (1987) and Pole and West (1990)), with all the robustness issues that surround them, are largely unnecessary.

The paper is organized as follows. The MDM is first set up in Section 2 and then

†*Address for correspondence*: Mathematics Department, University of Nottingham, University Park, Nottingham, NG7 2RD, UK.

0035–9246/93/55849

fully defined in Section 3. Results presented in the latter section are formally proved in Section 4. Section 5 examines the properties of two interesting special cases of MDMs and it is shown that, despite being non-linear, the first two moments of their joint forecast distributions can be calculated algebraically in closed form. Section 5.1 illustrates how an MDM has been used to forecast a brand's price relative to the rest of the market together with its market share and the form of the non-linearity that they exhibit is studied. Finally, Section 6 offers a discussion of some of the properties of an MDM. It is shown how an MDM, represented by a given graph of an influence diagram (defined later), exhibits a specific conditional 'Granger causal' structure. It is illustrated how such causal structures can be tested by using multiprocess modelling techniques.

Throughout the paper the following results from the theory on conditional independence and graphs will be used. Any random vectors **X**, **Y**, **Z** and **W** will have the following conditional independence properties:

(a) $\mathbf{X} \amalg \mathbf{Y} \mid (\mathbf{Y}, \mathbf{Z})$,

(b) $\mathbf{X} \amalg \mathbf{Y} \mid \mathbf{Z} \Leftrightarrow \mathbf{Y} \amalg \mathbf{X} \mid \mathbf{Z}$,

(c) $\mathbf{X} \amalg (\mathbf{Y}, \mathbf{Z}) \mid \mathbf{W} \Leftrightarrow \begin{cases} \mathbf{X} \amalg \mathbf{Y} \mid (\mathbf{Z}, \mathbf{W}) \\ \text{together with} \\ \quad \mathbf{X} \amalg \mathbf{Z} \mid \mathbf{W} \end{cases}$

where $\mathbf{X} \amalg \mathbf{Y} \mid \mathbf{Z}$ reads **X** is independent of **Y** given **Z** (see Dawid (1979)).

Let $(\mathbf{X}_1, \ldots, \mathbf{X}_m)$ be an ordered set of random vectors on which $\cdot \amalg \cdot \mid \cdot$ is defined. Suppose that the following $m-1$ conditional independence statements are given:

$$\mathbf{X}_r \amalg P'(\mathbf{X}_r) \mid P(\mathbf{X}_r) \qquad 2 \leqslant r \leqslant m \tag{1.1}$$

where $P(\mathbf{X}_r) \subseteq \{\mathbf{X}_1, \ldots, \mathbf{X}_{r-1}\}$ and $P'(\mathbf{X}_r) = \{\mathbf{X}_1, \ldots, \mathbf{X}_{r-1}\} \setminus P(\mathbf{X}_r)$ where, for any two sets A and B, $B \setminus A$ denotes those elements in B which are not in A. A set of conditional independence statements such as these can be usefully represented graphically. A graph $G(\mathbf{X}, \mathbf{E})$ consists of a set of *nodes* **X** and *edges* **E**, where a directed edge from X_j to X_k is denoted by $(X_j, X_k) \in \mathbf{E}$. A directed graph with a directed edge from $\mathbf{X}_i$ to $\mathbf{X}_r$ if and only if $\mathbf{X}_i \in P(\mathbf{X}_r)$ is called the *graph of an influence diagram*—see Howard and Matheson (1981), Shachter (1986) and Smith (1989, 1990). The set $P(\mathbf{X}_r)$ is known as the *parent set* of $\mathbf{X}_r$. A *path* of length m from $\mathbf{X}_i$ to $\mathbf{X}_k$ is a sequence of nodes, $\mathbf{X}_i = \mathbf{X}_{i,0}, \ldots, \mathbf{X}_{i,m} = \mathbf{X}_k$, where $(\mathbf{X}_{i,(j-1)}, \mathbf{X}_{i,j}) \in E$ for each $j = 1, \ldots, m$. The set of vertices with paths to $\mathbf{X}_k$ is called the *ancestor set* of $\mathbf{X}_k$ and is denoted by $\mathrm{an}(\mathbf{X}_k)$. The graph of an influence diagram together with the conditional independence statements (1.1) is called an *influence diagram*.

2. SETTING UP THE MODEL

This section describes how the heuristic causal relationships between component processes of a multivariate time series $\{\mathbf{Y}_k\}_{k \leqslant t}$ can define an appropriate forecasting model which forms the basis of the graphical MDM.

Causation between *processes* has been studied quite extensively in recent years. Granger's (1969) original definition of causality required that, for all time t, a best linear estimate of V_t based on $\{U_k\}_{k<t}$ and $\{V_k\}_{k<t}$ need only depend on $\{V_k\}_{k<t}$. A graphical representation of Granger causality is given by Lauritzen (1989). Here,

however, the stronger definition of Florens and Mouchart (1985) is used—they say that U is a *non-cause* of V, if for all time t

$$V_t \amalg \{U_k\}_{k<t} | \{V_k\}_{k<t}.$$

Thus the original idea of Granger is then replaced with a definition based on conditional independence itself and now the best estimate (rather than the best linear estimate) of V_t based on $\{U_k\}_{k<t}$ and $\{V_k\}_{k<t}$ need only depend on $\{V_k\}_{k<t}$. By using this definition a set of conditional independence statements *related to causality* can be defined across processes. Although causality defined in terms of conditional independence does not quite capture what is usually meant by causality (Holland, 1986), non-causality of X and Y can often reasonably be considered as a consequence of there being a lack of causal relation from X to Y. Wermuth and Lauritzen (1990) *heuristically* argued that, when dealing with graphical models, variables which are hypothesized to be causally linked should be connected by directed edges consistent with the direction of causality. Thus a heuristic influence diagram can be created representing the conditional independence *related to causality* between the variables at a fixed time period t.

For example, consider three brands $\{X, Y, Z\}$ in a business market so that the sales of brand X can be considered as a causal factor in determining Y's sales and the sales of brand Y can be considered as a causal factor in determining Z's sales. Let X_t, Y_t and Z_t represent the sales at time t of brands X, Y and Z respectively. The heuristic graph of the influence diagram consistent with the causal relationships in this market is given in Fig. 1.

Suppose that the same influence diagram always represents the variables at all time points so that $P\{Y_t\} = X_t$ and $P\{Z_t\} = Y_t$ for $t \geqslant 1$. Suppose further that the processes over all time points up to and including time t can be represented by the graph of the influence diagram of Fig. 2 such that

$$P\{X_t\} \subseteq \{\mathbf{X}^{t-1}\},$$
$$P\{Y_t\} \subseteq \{\mathbf{X}^{t}, \mathbf{Y}^{t-1}\},$$
$$P\{Z_t\} \subseteq \{\mathbf{Y}^{t}, \mathbf{Z}^{t-1}\}$$

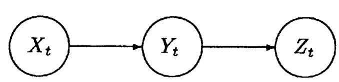

Fig. 1. Heuristic graph of the influence diagram representing the causal relationships between the sales at time t of brands X, Y and Z

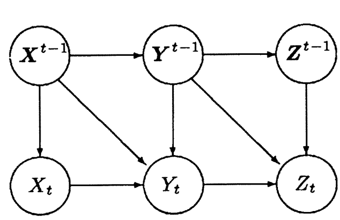

Fig. 2. Graph of the influence diagram representing the causal relationships between the processes $\{X_k\}_{k\leqslant t}$, $\{Y_k\}_{k\leqslant t}$ and $\{Z_k\}_{k\leqslant t}$

where $\mathbf{W}^t = \{W_1, \ldots, W_t\}^{\mathrm{T}}$ for any time series $\{W_k\}_{k \geqslant 1}$ such that A^{T} denotes the transpose of any matrix A.

It is clear from Fig. 2 that the conditional independences *related to causality* are such that

$$Y_t \amalg \{Z_k\}_{k<t} | \{Y_k\}_{k<t}, \{X_k\}_{k \leqslant t}.$$

This leads naturally to a general definition of *conditional causality* which extends Florens and Mouchart's (1985) definition so that U is a *non-cause* of V *given* W if for all points in time t

$$V_t \amalg \{U_k\}_{k<t} | \{V_k\}_{k<t}, \{W_k\}_{k \leqslant t}.$$

A best estimate of V_t not only now depends on the past series $\{V_k\}_{k<t}$ and $\{W_k\}_{k<t}$ but also on the current value W_t. So, in particular, Fig. 2 implies that Z is a non-cause of Y given X. Thus the best estimate of Y_t based on $\{X_k\}_{k \leqslant t}$, $\{Y_k\}_{k<t}$ and $\{Z_k\}_{k \leqslant t}$ need only depend on $\{Y_k\}_{k<t}$ and $\{X_k\}_{k \leqslant t}$. An appropriate forecasting model for each variable at time t is simply a function of its own past series, the past series of its parents and the value of its parents *at time t*. For example, the general form of an appropriate forecasting model for the series $\{Y_k\}_{k \geqslant t}$ is given by

$$Y_t = f(\mathbf{x}^t, \mathbf{y}^{t-1}, \boldsymbol{\theta}_t) + v_t,$$

where $f(\;)$ is some function and v_t has some probability distribution. This reasoning can be generalized directly to the case where there are n series.

The MDM uses this methodology for deriving its observation equations and through the special form of the system equation breaks a complex multivariate problem into n univariate problems. Notice how the causal relationships between the variables are accommodated and that the stringent symmetry conditions imposed on the variables in the models of Harvey (1986) and Quintana and West (1987) are not required. The MDM will now be formally defined.

3. MULTIREGRESSION DYNAMIC MODEL

Let $\mathbf{Y}_t^{\mathrm{T}} = \{Y_t(1), Y_t(2), \ldots, Y_t(n)\}$ denote the general n-dimensional multivariate time series at time t. The observed value of $Y_t(r)$ is denoted by $y_t(r)$; let $\mathbf{y}^t(r)^{\mathrm{T}} = \{y_1(r), \ldots, y_t(r)\}$ and set

$$\mathbf{X}_t(r)^{\mathrm{T}} = \{Y_t(1), Y_t(2), \ldots, Y_t(r-1)\} \qquad 2 \leqslant r \leqslant n,$$
$$\mathbf{Z}_t(r)^{\mathrm{T}} = \{Y_t(r+1), \ldots, Y_t(n)\} \qquad 2 \leqslant r \leqslant n-1.$$

Suppose that for any fixed time t the structure of the series $\mathbf{Y}_t$ can be heuristically represented by an influence diagram I such that

$$P\{Y_t(r)\} \subseteq \mathbf{X}_t(r).$$

Suppose further that the process $\{\mathbf{Y}_k\}_{k \leqslant t}$ can be heuristically represented by an influence diagram I^* such that

$$P\{Y_t(r)\} \subseteq \{\mathbf{X}^t(r), \mathbf{Y}^{t-1}(r)\}.$$

Let $\boldsymbol{\theta}_t^{\mathrm{T}} = \{\boldsymbol{\theta}_t(1)^{\mathrm{T}}, \boldsymbol{\theta}_t(2)^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_t(n)^{\mathrm{T}}\}$ be the state vectors determining the distributions of $Y_t(1), Y_t(2), \ldots, Y_t(n)$ respectively, and let s_r be the dimension of the vector $\boldsymbol{\theta}_t(r)$, $t \geqslant 1$. For notational convenience set

$$\boldsymbol{\phi}_t(r)^{\mathrm{T}} = \{\boldsymbol{\theta}_t(1)^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_t(r-1)^{\mathrm{T}}\} \qquad 2 \leqslant r \leqslant n,$$
$$\boldsymbol{\psi}_t(r)^{\mathrm{T}} = \{\boldsymbol{\theta}_t(r+1)^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_t(n)^{\mathrm{T}}\} \qquad 2 \leqslant r \leqslant n-1.$$

Call $\{\mathbf{Y}_t\}_{t \geqslant 1}$ an MDM if it is governed by the following n observation equations, system equation and initial information where the observation and system equations are conditional on the past and the restrictions on these equations given below hold for all points in time:

(a) *observation equations—*

$$Y_t(r) = \mathbf{F}_t(r)^{\mathrm{T}} \boldsymbol{\theta}_t(r) + v_t(r) \qquad v_t(r) \sim (0, V_t(r)), \quad 1 \leqslant r \leqslant n; \tag{3.1}$$

(b) *system equation—*

$$\boldsymbol{\theta}_t = G_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t \qquad \mathbf{w}_t \sim (0, W_t); \tag{3.2}$$

(c) *initial information—*

$$(\boldsymbol{\theta}_0 | D_0) \sim (\mathbf{m}_0, C_0). \tag{3.3}$$

The s_r-dimensional column vector $\mathbf{F}_t(r)$ is allowed to be an arbitrary but known function of $\mathbf{x}^t(r)$ and $\mathbf{y}^{t-1}(r)$, but not $\mathbf{z}^t(r)$ and $y_t(r)$; $V_t(1), \ldots, V_t(n)$ are the known scalar observation variances; the $s \times s$ matrices $G_t = \mathrm{blockdiag}\{G_t(1), \ldots, G_t(n)\}$, $W_t = \mathrm{blockdiag}\{W_t(1), \ldots, W_t(n)\}$ and $C_0 = \mathrm{blockdiag}\{C_0(1), \ldots, C_0(n)\}$ are assumed known and are such that $G_t(r)$, $W_t(r)$ and $C_0(r)$ are $s_r \times s_r$ square matrices which may be functions of past vectors $\mathbf{x}^{t-1}(r)$ and $\mathbf{y}^{t-1}(r)$ but nothing else. Thus $\{\mathbf{Y}_t(r) | \mathbf{y}^{t-1}, \mathbf{F}_t(r), \boldsymbol{\theta}_t(r)\}$ follows some distribution with mean $\mathbf{F}_t(r)^{\mathrm{T}} \boldsymbol{\theta}_t(r)$ and variance $V_t(r)$.

The error vectors $\mathbf{v}_t^{\mathrm{T}} = \{v_t(1), \ldots, v_t(n)\}$ and $\mathbf{w}_t^{\mathrm{T}} = \{\mathbf{w}_t(1)^{\mathrm{T}}, \ldots, \mathbf{w}_t(n)^{\mathrm{T}}\}$, where $v_t(r)$ is the observation error and $\mathbf{w}_t(r)$ is the s_r-dimensional system error vector for $Y_t(r)$, $1 \leqslant r \leqslant n$, are such that variables $v_t(1), \ldots, v_t(n)$ and $\mathbf{w}_t(1), \ldots, \mathbf{w}_t(n)$ are all mutually independent and the vectors $\{\mathbf{v}_t, \mathbf{w}_t\}_{t \geqslant 1}$ are mutually independent with time.

To illustrate how an MDM would be defined for a vector time series, consider the following example. Suppose that $\mathbf{Y}_t^{\mathrm{T}} = \{Y_t(1), \ldots, Y_t(4)\}$ where, at any fixed time t, $\mathbf{Y}_t$ can be represented by the graph of the influence diagram given in Fig. 3. The MDM for $\mathbf{Y}_t$ would have observation equations in which $\mathbf{F}_t(1)$ is a function of $\mathbf{y}^{t-1}(1)$; $\mathbf{F}_t(2)$ is a function of $\{\mathbf{y}^t(1), \mathbf{y}^{t-1}(2)\}$; $\mathbf{F}_t(3)$ is a function of $\{\mathbf{y}^t(1), \mathbf{y}^t(2), \mathbf{y}^{t-1}(3)\}$; $\mathbf{F}_t(4)$ is a function of $\{\mathbf{y}^t(2), \mathbf{y}^t(3), \mathbf{y}^{t-1}(4)\}$.

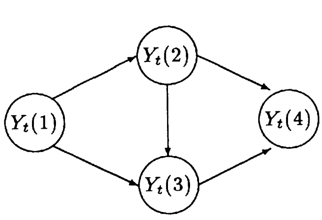

Fig. 3. Graph of the influence diagram for $Y_t(1), \ldots, Y_t(4)$

There are two important results which are central to the theory of MDMs, the proofs of which will be given in the next section. It is shown that if

$$\amalg_{r=1}^{n} \boldsymbol{\theta}_{t-1}(r) \mid \mathbf{y}^{t-1}$$

then the following conditional independence statements must hold.

Result 1.

$$\amalg_{r=1}^{n} \boldsymbol{\theta}_{t}(r) \mid \mathbf{y}^{t}. \tag{3.4}$$

In words this means that if $\{\boldsymbol{\theta}_{t-1}(r)\}$ are mutually independent given the data $\mathbf{y}^{t-1}$ then under the DLM $\{\boldsymbol{\theta}_t(r)\}$ are also independent given $\mathbf{y}_t$. It will follow by induction that, provided that $\{\boldsymbol{\theta}_0(r)\}$ are initially independent, the parameters remain independent for all time given the current available information.

Result 2.

$$\boldsymbol{\theta}_t(r) \amalg \mathbf{z}^t(r) \mid \mathbf{x}^t(r), \mathbf{y}^t(r). \tag{3.5}$$

Written in terms of the components of $\mathbf{y}^t$ this reads

$$\boldsymbol{\theta}_t(r) \amalg \mathbf{y}^t(r+1), \mathbf{y}^t(r+2), \ldots, \mathbf{y}^t(n) \mid \mathbf{y}^t(1), \ldots, \mathbf{y}^t(r).$$

In words this means that, given the past observation vectors of the first r indexed series, $\boldsymbol{\theta}_t(r)$ is independent of the rest of the past data.

When defining the MDM, C_0 is set to be block diagonal and so the parameters for each variable are initially mutually independent. Therefore, by result 1, the parameters associated with each variable are updated independently after each observation and remain independent at each time point. Thus, as each variable follows a conditional univariate Bayesian dynamic model, the conditional distribution for each variable can be updated independently and conditional forecasts can be found separately. Bayesian forecasting techniques such as intervention, trends and seasonals can be directly transferred on to these models. A complex multivariate problem has therefore been decomposed into n univariate problems.

No assumption of normality has been made for either results 1 and 2 or when defining the MDM. It follows that large classes of multivariate processes such as those discussed by Kitagawa (1987) can be used as the components of the decomposition. However, even when $\mathbf{F}_t(r)$ is a non-linear function of $\{\mathbf{x}^t(r), \mathbf{y}^{t-1}(r)\}$, because it is assumed known and therefore fixed at time t, a *normal MDM* can be defined such that $\mathbf{v}_t$ and $\mathbf{w}_t$ are Gaussian so that, *conditional on* $\mathbf{x}_t(r)$, $Y_t(r)$ is Gaussian. The MDM is particularly simple to work with in this case. Each variable follows a normal DLM and, as such, updating and forecasts of conditional univariate distributions are identical with those given by Harrison and Stevens (1976). Section 5 discusses two special cases of normal MDMs called the *linear multiregression dynamic model* (LMDM) and the *corrected linear multiregression dynamic model* (CLMDM) in which $\mathbf{F}_t(r)$ is a linear function of $\{\mathbf{x}^t(r), \mathbf{y}^{t-1}(r)\}$.

In reality, the series are observed simultaneously so that $\mathbf{x}_t(r)$ will not be observed before a forecast for $Y_t(r)$ is made. The marginal forecast distributions for each variable are therefore required. In general these marginal distributions will not be Gaussian. However, the moments of $\mathbf{Y}_t$ for many MDMs can be derived fairly easily. The derivation of the first two moments of $\mathbf{Y}_t$ for the CLMDM and the first moment of the LMDM are illustrated in Section 5.

Results 1 and 2 combine to enable the joint one-step-ahead forecast distribution of $\mathbf{Y}_t$ to be simplified. Consider the forecast distribution for n brands:

$$p(\mathbf{y}_t|\mathbf{y}^{t-1}) = \int_{\boldsymbol{\theta}_t} p\{\mathbf{y}_t|\boldsymbol{\theta}_t, \mathbf{y}^{t-1}\}\, p\{\boldsymbol{\theta}_t|\mathbf{y}^{t-1}\}\, \mathrm{d}\boldsymbol{\theta}_t.$$

The conditional independence statements of result 1, together with the structure of the vector $\mathbf{F}_t(r)$ specified by the MDM, ensure that the joint distribution can be expressed as the product of the individual forecast distributions of $y_t(r)$ with regressors $\mathbf{x}_t(r)$. So

$$\begin{aligned} p(\mathbf{y}_t|\mathbf{y}^{t-1}) &= \prod_r p\{y_t(r)|\mathbf{x}^t(r), \mathbf{y}^{t-1}(r)\} \\ &= \prod_r \int_{\boldsymbol{\theta}_t(r)} p\{y_t(r)|\mathbf{x}^t(r), \mathbf{y}^{t-1}(r), \boldsymbol{\theta}_t(r)\}\, p\{\boldsymbol{\theta}_t(r)|\mathbf{y}^{t-1}\}\, \mathrm{d}\boldsymbol{\theta}_t(r). \end{aligned}$$

However, by the conditional independence statements of result 2 and the specified structure of $G_t(r)$, $p(\boldsymbol{\theta}_t(r)|\mathbf{y}^{t-1})$ only depends on $\mathbf{x}^{t-1}(r)$ and $\mathbf{y}^{t-1}(r)$ and thus can be rewritten as

$$p\{\boldsymbol{\theta}_t(r)|\mathbf{y}^{t-1}\} = p\{\boldsymbol{\theta}_t(r)|\mathbf{x}^{t-1}(r), \mathbf{y}^{t-1}(r)\}.$$

The next section formally proves results 1 and 2.

4. FORMAL PRESENTATION OF RESULTS

To prove the results of the last section, the following theorem given by Pearl and Verma (1987), Pearl (1988) and in its present form by Lauritzen *et al.* (1990) is required. It identifies all the conditional independences defined by an influence diagram and gives an algorithm for deciding whether any given conditional independences can be logically deduced from an influence diagram.

Theorem 1. Suppose that conditional independence statements for a set of variables are represented in an influence diagram, whose graph is I, and that U, V and W denote sets of variables on it. Adapt the influence diagram in the following way.

(a) Form the directed subgraph I_1 of I whose nodes are in the ancestor set an(U, V, W) and whose directed edges are those in I which lie between these nodes.

(b) Join all pairs of nodes $(X, Y) \in P(Z)$ by an undirected edge, where $P(Z)$ is the parent set of Z and $Z \in I_1$. This process is known as moralizing the graph since all parents of a single variable are joined by an arc. Call this mixed graph I_2.

(c) Form an undirected graph J by replacing all directed edges in I_2 by undirected edges.

Then

$$U \amalg V | W$$

if all undirected paths in J between a node $R \in U$ and $S \in V$ must pass through a node $T \in W$.

Furthermore, if this condition is violated, then there is a probabilistic influence diagram respecting the conditional independence statements of the influence diagram for which $U \amalg V | W$ is not true.

Theorem 1 can now be directly used to prove the following theorem.

Theorem 2. Let $\{\mathbf{Y}_t\}_{t \geq 1}$ be governed by an MDM and, using the notation of the previous section, assume that

$$\boldsymbol{\phi}_{t-1}(r) \amalg \mathbf{y}^{t-1}(r), \mathbf{z}^{t-1}(r) \,|\, \mathbf{x}^{t-1}(r) \qquad r=2, \ldots, n, \tag{4.1}$$

$$\boldsymbol{\theta}_{t-1}(r) \amalg \mathbf{z}^{t-1}(r), \boldsymbol{\phi}_{t-1}(r) \,|\, \mathbf{x}^{t-1}(r), \mathbf{y}^{t-1} \qquad r=1, \ldots, n, \tag{4.2}$$

$$\boldsymbol{\psi}_{t-1}(r) \amalg \boldsymbol{\theta}_{t-1}(r), \boldsymbol{\phi}_{t-1}(r) \,|\, \mathbf{y}^{t-1} \qquad r=2, \ldots, n-1. \tag{4.3}$$

Then the following conditional independence statements must also be true:

$$\boldsymbol{\phi}_{t}(r) \amalg \mathbf{y}^{t}(r), \mathbf{z}^{t}(r) \,|\, \mathbf{x}^{t}(r) \qquad r=2, \ldots, n, \tag{4.4}$$

$$\boldsymbol{\theta}_{t}(r) \amalg \mathbf{z}^{t}(r), \boldsymbol{\phi}_{t}(r) \,|\, \mathbf{x}^{t}(r), \mathbf{y}^{t}(r) \qquad r=1, \ldots, n, \tag{4.5}$$

$$\boldsymbol{\psi}_{t}(r) \amalg \boldsymbol{\theta}_{t}(r), \boldsymbol{\phi}_{t}(r) \,|\, \mathbf{y}^{t} \qquad r=2, \ldots, n-1. \tag{4.6}$$

Proof. To prove this result, it is first necessary that the conditional independence statements contained in the inductive hypotheses (4.1)–(4.3) and the MDM itself are represented graphically. By using theorem 1 it is then simple to show that conditional independence statements (4.4)–(4.6) must also hold.

For clarity of presentation, the inductive hypotheses (4.1), (4.2) and (4.3) are represented by the *directed arcs* in the separate graphs of influence diagrams of Figs 4, 5 and 6 respectively. In each diagram, variables have been allowed to share nodes when it is convenient to do so and when there is no loss of information. Thus, for example, in Fig. 4 $\mathbf{z}^{t-1}(r-1) = \{\mathbf{y}^{t-1}(r), \mathbf{z}^{t-1}(r)\}$ and $\boldsymbol{\psi}_{t-1}(r-1) = \{\boldsymbol{\theta}_{t-1}(r), \boldsymbol{\psi}_{t-1}(r)\}$. The justification for the missing directed arcs in these graphs of influence diagrams will now be presented (for the moment all undirected arcs are ignored).

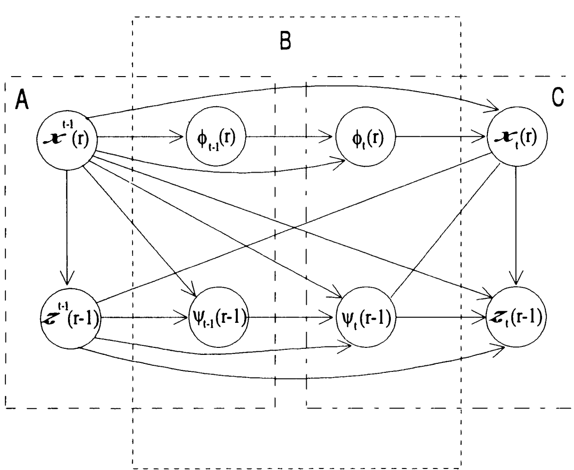

Fig. 4. The directed arcs of this graph give the graph of the influence diagram for inductive hypothesis (4.1); the directed and undirected arcs make up the moralized graph

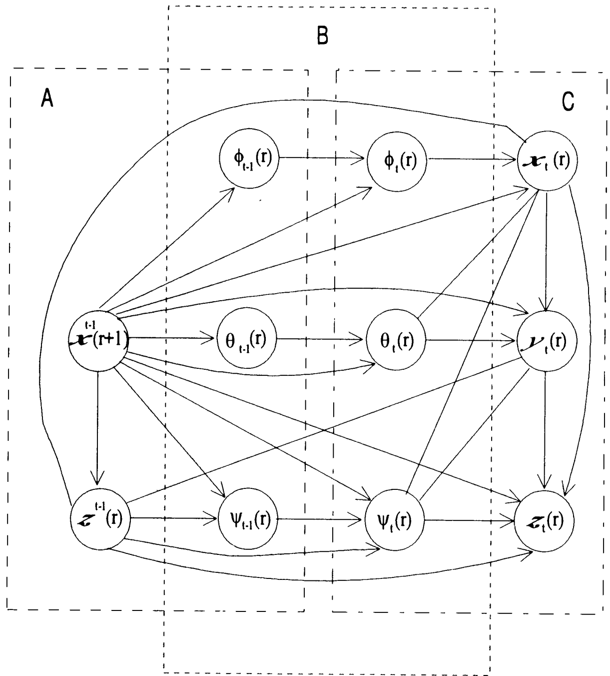

Fig. 5. The directed arcs of this graph give the graph of the influence diagram for inductive hypothesis (4.2); the directed and undirected arcs make up the moralized graph

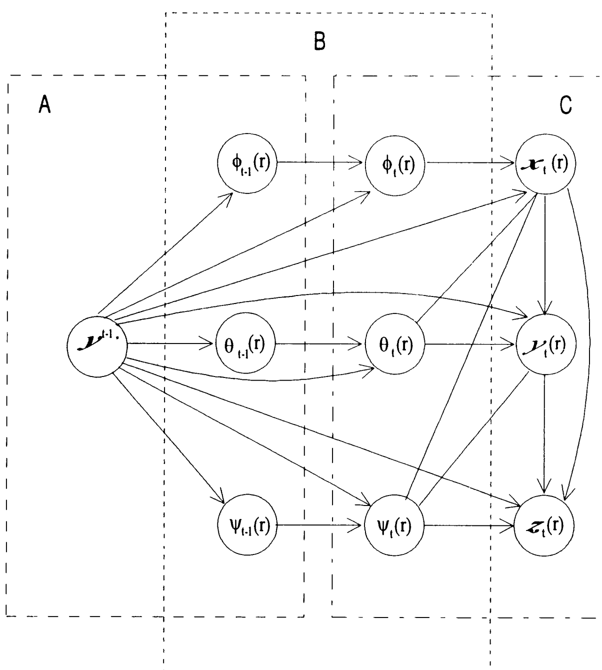

Fig. 6. The directed arcs of this graph give the graph of the influence diagram for inductive hypothesis (4.3); the directed and undirected arcs make up the moralized graph

Using property (c) of Section 1, assumption (4.2) can be equivalently stated by the pair of assertions

$$\boldsymbol{\theta}_{t-1}(r) \amalg \boldsymbol{\phi}_{t-1}(r) \mid \mathbf{y}^{t-1}, \tag{4.7}$$

$$\boldsymbol{\theta}_{t-1}(r) \amalg \mathbf{z}^{t-1}(r) \mid \mathbf{x}^{t-1}(r), \mathbf{y}^{t-1}(r). \tag{4.8}$$

Now, by assumption (4.1) and assertion (4.8), arcs between nodes of subvectors $\mathbf{y}^{t-1}$ whose components all have an index in $\mathbf{y}_t$ greater than r to those nodes of subvectors of $\boldsymbol{\theta}_t$ whose components have indices less than or equal to r are allowed to be omitted. In contrast, assertion (4.7) and assumption (4.3) allow arcs between parameter vectors which do not share the same components to be omitted. This justifies the implied conditional independence statements in the boxes labelled A in Figs 4–6.

The block diagonal forms of G_t and W_t in the MDM ensure that

$$\boldsymbol{\theta}_t(r) \amalg \boldsymbol{\theta}_{t-1} \setminus \{\boldsymbol{\theta}_{t-1}(r)\} \mid \mathbf{y}^{t-1}, \boldsymbol{\theta}_{t-1}(r) \qquad 1 \leqslant r \leqslant n.$$

Thus on the graph of the influence diagram any arcs between sets of components of $\boldsymbol{\theta}_{t-1}$ and sets of components of $\boldsymbol{\theta}_t$ with no common index can be omitted. This argument justifies the omission of arcs in box B of the influence diagram boxes.

Finally the conditional independence statements implicit in the observation equations of the MDM mean that

$$y_t(r) \amalg \boldsymbol{\theta}_t \setminus \{\boldsymbol{\theta}_t(r)\} \mid \mathbf{y}^{t-1}(r), \mathbf{x}_t(r), \boldsymbol{\theta}_t(r).$$

These statements justify the omission of arcs in box C of the influence diagram graphs.

The undirected arcs in Figs 4–6 are those extra arcs which are introduced when each of the influence diagrams in moralized. It is now simple to check that the set of nodes representing the conditioning variables block all paths between the nodes associated with elements in different sets U and V claimed to be independent. So by theorem 1 the result is proved. □

The results introduced in Section 3 can now be formally stated in the following corollary of theorem 2.

Corollary 1. If in an MDM the initial states are independent, i.e. if $\amalg_{r=1}^{n} \boldsymbol{\theta}_0(r)$, then for all times t

$$\amalg_{r=1}^{n} \boldsymbol{\theta}_t(r) \mid \mathbf{y}^t$$

and

$$\boldsymbol{\theta}_t(r) \amalg \mathbf{y}^t(r+1), \ldots, \mathbf{y}^t(n) \mid \mathbf{y}^t(1), \ldots, \mathbf{y}^t(r).$$

Proof. To prove the result for $t=1$ proceed as in theorem 2. From the hypotheses (set A), the system equation (set B) and the observation equations (set C) the graph of the influence diagram I is represented by the *directed arcs* of the graph in Fig. 7.

By using exactly the same argument as in theorem 2, again, the undirected arcs of Fig. 7 are those introduced when the graph is moralized, and so the conditional independence statements (4.4)–(4.6) of theorem 2 can now be deduced for time $t=1$. Therefore, since $\amalg_{r=1}^{n} \boldsymbol{\theta}_0(t)$ under the corollary, theorem 2 must be true for $t=1$, and so by induction the assertions of theorem 2 must be true for all times t.

Since theorem 2 holds for all times t, the conditional independence statements from the theorem can be combined:

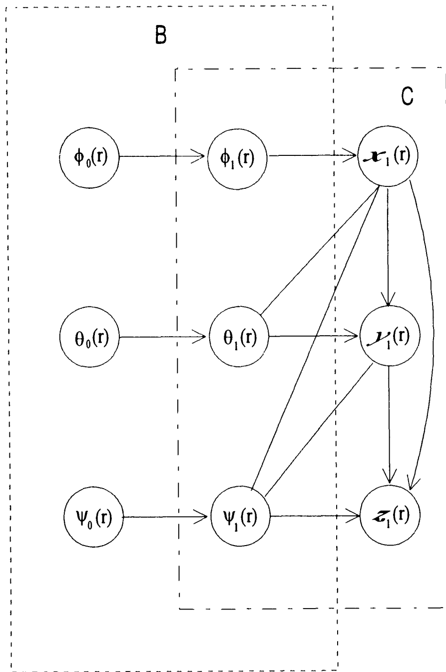

Fig. 7. The directed arcs of this graph give the graphs of the influence diagram for the proof of corollary 1; the directed and undirected arcs make up the moralized graph

$$\boldsymbol{\theta}_t(r) \amalg \{\boldsymbol{\phi}_t(r), \boldsymbol{\psi}_t(r)\}, \mathbf{z}^t(r) \,|\, \mathbf{x}^t(r), \mathbf{y}^t(r). \tag{4.9}$$

Now, by property (c) (Section 1) statement 4.9 implies that

$$\boldsymbol{\theta}_t(r) \amalg \{\boldsymbol{\phi}_t(r), \boldsymbol{\psi}_t(r)\} \,|\, \mathbf{z}^{t-1}(r), \mathbf{x}^{t-1}(r), \mathbf{y}^{t-1}(r);$$

therefore,

$$\amalg_{r=1}^{n} \boldsymbol{\theta}_t(r) \,|\, \mathbf{y}^t.$$

Statement (4.9) also implies that

$$\boldsymbol{\theta}_t(r) \amalg \mathbf{z}^t(r) \,|\, \mathbf{x}^t(r), \mathbf{y}^t(r)$$

and so

$$\boldsymbol{\theta}_t(r) \amalg \mathbf{y}^t(r+1), \ldots, \mathbf{y}^t(n) \,|\, \mathbf{y}^t(1), \ldots, \mathbf{y}^t(r);$$

therefore the corollary and hence the results of Section 3 are proved. □

These results apply to a very wide class of models. However, it is helpful initially to consider the implications for the LMDM and CLMDM because the joint distribution of the variables in these models can be calculated explicitly and the one-step-ahead forecasts have a particularly simple form.

5. LINEAR MULTIREGRESSION DYNAMIC MODELS

Consider the MDM defined by equations (3.1)–(3.3). This section introduces two special cases of the MDM—the LMDM and the CLMDM—which are especially simple to work with.

Suppose that $\{v_t(r), \mathbf{w}_t(r)\}_{t\geqslant 1}$ are jointly Gaussian, are independent of $\mathbf{y}^{t-1}$ and $\mathbf{F}_t(1)$ does not depend on $\mathbf{Y}_t$. Now, let

$$\mathbf{F}_t(r)^{\mathrm{T}} = \{\mathbf{x}_t(r)^{\mathrm{T}}, \tilde{\mathbf{x}}_t(r)^{\mathrm{T}}\}, \qquad 2 \leqslant r \leqslant n,$$

where $\mathbf{x}_t(r)^{\mathrm{T}} = \{y_t(1), \ldots, y_t(r-1)\}$ and $\tilde{\mathbf{x}}_t(r)$ is a set of known exogenous variables that are not dependent on $\mathbf{x}_t(r)$. This process is an LMDM. Alternatively, suppose that

$$\mathbf{F}_t(r)^{\mathrm{T}} = [\{\mathbf{x}_t(r) - \hat{\mathbf{f}}_t(r)\}^{\mathrm{T}}, \tilde{\mathbf{x}}_t(r)^{\mathrm{T}}],$$

where

$$\hat{\mathbf{f}}_t(r)^{\mathrm{T}} = [E\{Y_t(1)|\mathbf{y}^{t-1}\}, \ldots, E\{Y_t(r-1)|\mathbf{y}^{t-1}\}] \tag{5.1}$$

and, from result 2 of Section 3, $\hat{\mathbf{f}}_t(r)$ only depends on $\mathbf{y}^{t-1}$ through $\mathbf{x}^{t-1}(r)$ and $\mathbf{y}^{t-1}(r)$. Once again $\tilde{\mathbf{x}}_t(r)$ is a set of known exogenous variables that are not dependent on $\mathbf{x}_t(r)$. This process is a CLMDM. Thus, given the components $\mathbf{w}_t(r)$ which precede it, each component $Y_t(r)$ is described as a *univariate* DLM with regressors contained in the vectors $\mathbf{F}_t(r)$ given above.

Note that the LMDM is a stochastic version of a recursive simultaneous equations model where $Y_t(r)$ is regressed against a subset of contemporary variables listed before $Y_t(r)$. Harvey (1989) considers a degenerate special case of this. Unlike in economics, however, the emphasis here is not on parameter estimation but on the types of predictive joint distribution that $\mathbf{Y}_t$ can exhibit given the inevitable continued uncertainty about the regression state in the process.

The causal relationships between brands modelled by an LMDM have a different interpretation from those modelled by a CLMDM. For example, suppose that there are just two variables, $\{Y_k(1)\}_{k\geqslant 1}$ and $\{Y_k(2)\}_{k\geqslant 1}$, such that $Y(1)$ is thought to be a causal factor of $Y(2)$. If $\{\mathbf{Y}_k\}_{k\geqslant 1}$ follows an LMDM, then a long-term change in $Y(1)$'s level would cause a sustained level change in $Y(2)$. However, if $\{\mathbf{Y}_k\}_{k\geqslant 1}$ follows a CLMDM, then the same long-term level change in $Y(1)$ would have a different causal effect on $Y(2)$. In this case the level change in $Y(2)$ would be short term, followed by a drift back to the original level. This is due to the fact that the CLMDM uses residuals as regressors unlike the LMDM which uses the actual observation. Thus the CLMDM essentially relates causality through forecast residuals so that misforecasting of $Y_t(1)$ helps in adjusting the forecast distribution of $Y_t(2)$ whereas the reverse is not true. Therefore, a sudden level change in $Y(1)$ would lead to a large residual value in the model for $Y(2)$, thus leading to a level change in $Y(2)$. As the model for $Y(1)$ adapts to the level change, so the residual in the regression term in $Y(2)$'s model will become smaller and $Y(2)$ will drift back to its original level. To illustrate a situation in which a CLMDM would be a useful model, consider an ice-cream market. Suppose that $Y_t(1)$ is the total sales of ice-cream and $Y_t(2)$ is the market share of ice-cream type A. Now if there were a heat wave the total demand $Y_t(1)$ of ice-cream might suddenly increase. If brand A had extra stock and could cope with the extra demand better than another brand B, say, then the market share $Y_t(2)$ would be expected to increase over the heat wave. This situation could be modelled by a CLMDM such that

$$Y_t(2) = \theta_t + \beta_t\{Y_t(1) - E[Y_t(1)|\mathbf{y}^{t-1}]\} + \epsilon_t,$$

where ϵ_t is some error term and β_t is some parameter associated with A's ability to cope with extra demand.

For both the LMDM and the CLMDM, the separate components $\{Y_t(r)|\mathbf{x}_t(r)\}$ have either a Gaussian or Student t-distribution depending on whether the forecast distribution is assumed known or estimated. The forecast distributions and updating relationships associated with these conditional univariate models are identical with those of West and Harrison (1989). It then follows from theorem 2 that the one-step-ahead forecast density of $\mathbf{Y}_t$ is simply the product of the univariate conditional one-step-ahead forecast densities of $\{Y_t(r)|\mathbf{x}_t(r)\}$, for $r=1, \ldots, n$. Although the joint forecast distribution of $\mathbf{Y}_t$ will not, in general, be Gaussian, its mean and covariance matrix take a relatively simple form and these will now be derived here. Assume throughout this derivation that all means and variances are found conditionally on $\tilde{\mathbf{x}}_t(r)$.

The mean of the marginal forecast distribution for $\mathbf{Y}_t$ when it follows an LMDM will firstly be derived. As has already been mentioned, each variable under the LMDM follows a univariate regression DLM. Using the forecast distributions for the regression DLMs given by West and Harrison (1989), p. 111, it is clear that the expectation of the conditional forecast distribution of $\{Y_t(r)|\mathbf{x}_t(r)\}$ given the past can then be rewritten as

$$\begin{aligned} E\{Y_t(1)|\mathbf{y}^{t-1}(1)\} &= a_t^{(0)}(1), \\ E\{Y_t(r)|\mathbf{y}^{t-1}(r), \mathbf{x}^t(r)\} &= a_t^{(0)}(r) + \sum_{i=1}^{r-1} a_t^{(i)}(r)\, Y_t(i) \qquad 2 \leqslant r \leqslant n \end{aligned} \tag{5.2}$$

where $a_t^{(0)}(r)$ is a function of $\{\tilde{\mathbf{x}}_t(r), \mathbf{y}^{t-1}(r), \mathbf{x}^{t-1}(r)\}$ only.

The marginal forecast means for $\mathbf{Y}_t$ given the past can then be easily calculated from these conditional means by using the identity

$$E\{Y_t(r)|\mathbf{y}^{t-1}(r), \mathbf{x}^{t-1}(r)\} = E[E\{Y_t(r)|\mathbf{y}^{t-1}(r), \mathbf{x}^t(r)\}]. \tag{5.3}$$

Write $\mathbf{a}_t^{(0)\mathrm{T}} = \{a_t^{(0)}(1), \ldots, a_t^{(0)}(n)\}$ and A as the $n \times n$ lower triangular matrix whose $\{j, k)$th element a_{jk} is given by

$$a_{jk} = \begin{cases} a_t^{(k)}(j) & k < j, \\ 0 & \text{otherwise.} \end{cases}$$

Then, by using identity (5.3), equation (5.2) across all $r=1, \ldots, n$ can be expressed by

$$E(\mathbf{Y}_t|\mathbf{y}^{t-1}) = \mathbf{a}_t^{(0)} + A\, E(\mathbf{Y}_t|\mathbf{y}^{t-1})$$

and so

$$E(\mathbf{Y}_t|\mathbf{y}^{t-1}) = (I-A)^{-1}\mathbf{a}_t^{(0)}.$$

Similarly, for the CLMDM, the expectation of the forecast distribution for $\{Y_t(r)|\mathbf{x}_t(r)\}$ given the past can be rewritten as

$$\begin{aligned} E\{Y_t(1)|\mathbf{y}^{t-1}(1)\} &= \bar{a}_t^{(0)}(1), \\ E\{Y_t(r)|\mathbf{y}^{t-1}(r), \mathbf{x}^t(r)\} &= \bar{a}_t^{(0)}(r) + \sum_{t=1}^{r-1} \bar{a}_t^{(i)}(r)\{Y_t(i) - \hat{\mathbf{f}}_t(i)\} \qquad 2 \leqslant r \leqslant n \end{aligned} \tag{5.4}$$

where $\bar{a}_t^{(j)}(r)$ corresponds to a different value of the analogous $a_t^{(j)}(r)$ of equation (5.2) and $\hat{\mathbf{f}}_t(r)$ is the same as in equation (5.1). By using identity (5.3) it is obvious that the marginal forecast means for $\mathbf{Y}_t(r)$ following the CLMDM are simply

$$E\{Y_t(r)\,|\,\mathbf{y}^{t-1}(r), \mathbf{x}^{t-1}(r)\} = \bar{a}_t^{(0)}(r).$$

Although the marginal covariance matrix of $\mathbf{Y}_t$ given the past is found from the same recursive relationships in both the LMDM and the CLMDM, the covariance matrix of the CLMDM takes a simpler form and so only the derivation of this matrix will be shown here. To find this matrix it is first necessary to find the variance of the marginal forecast distribution of $Y_t(r)$, for $r=1, \ldots, n$.

Suppose that $\boldsymbol{\theta}_t(r)^{\mathrm{T}} = \{\boldsymbol{\theta}_t^*(r)^{\mathrm{T}}, \tilde{\boldsymbol{\theta}}_t(r)^{\mathrm{T}}\}$ where $\boldsymbol{\theta}_t^*(r)$ is the set of parameters contained in $\boldsymbol{\theta}_t(r)$ associated with $\mathbf{x}_t(r)$ and $\tilde{\boldsymbol{\theta}}_t(r)$ is the set associated with $\tilde{\mathbf{x}}_t(r)$. If $R_t(r)$ is the covariance matrix of the prior distribution of $\{\boldsymbol{\theta}_t(r)\,|\,\mathbf{y}^{t-1}\}$ then $R_t(r)$ can be expressed by

$$R_t(r) = \begin{pmatrix} R_t^*(r) & R_t'(r) \\ R_t'(r)^{\mathrm{T}} & \tilde{R}_t(r) \end{pmatrix}$$

where

$$\begin{aligned} R_t^*(r) &= \operatorname{cov}\{\boldsymbol{\theta}_t^*(r), \boldsymbol{\theta}_t^*(r)\,|\,\mathbf{y}^{t-1}\}, \\ \tilde{R}_t(r) &= \operatorname{cov}\{\tilde{\boldsymbol{\theta}}_t(r), \tilde{\boldsymbol{\theta}}_t(r)\,|\,\mathbf{y}^{t-1}\}, \\ R_t'(r) &= \operatorname{cov}\{\boldsymbol{\theta}_t^*(r), \tilde{\boldsymbol{\theta}}_t(r)\,|\,\mathbf{y}^{t-1}\}. \end{aligned}$$

Therefore, the variance of the conditional forecast distribution of $\{Y_t(r)\,|\,\mathbf{x}_t(r)\}$ given by West and Harrison (1989), p. 111, can be rewritten as

$$\begin{aligned} \operatorname{var}\{Y_t(1)\,|\,\mathbf{y}^{t-1}(1)\} &= \tau_t^2(1), \\ \operatorname{var}\{Y_t(r)\,|\,\mathbf{y}^{t-1}(r), \mathbf{x}^t(r)\} &= \tau_t^2(r) + h\{\mathbf{x}_t(r)\} \qquad 2 \leqslant r \leqslant n \end{aligned} \tag{5.5}$$

where

$$h\{\mathbf{x}_t(r)\} = \operatorname{trace}[\mathbf{S}_t(r)^{\mathrm{T}}\{\mathbf{x}_t(r) - \hat{\mathbf{f}}_t(r)\}\{\mathbf{x}_t(r) - \hat{\mathbf{f}}_t(r)\}^{\mathrm{T}} S_t(r)]$$

where each scalar $\tau_t^2(r) \geqslant 0, r=1, \ldots, n$, is a function of $\{\tilde{\mathbf{x}}_t(r), \mathbf{y}^{t-1}(r), \mathbf{x}^{t-1}(r), \tilde{R}_t(r), R_t'(r), V_t(r)\}$ and $S_t(r)\,S_t(r)^{\mathrm{T}} = R_t^*(r)$.

To find the variance of the marginal forecast distribution of $\mathbf{Y}_t(r)$ given the past, it is first noted that for any two variables $Z(1)$ and $Z(2)$ the following identity holds:

$$\operatorname{var}\{Z(2)\} = E[\operatorname{var}\{Z(2)\,|\,Z(1)\}] + \operatorname{var}[E\{Z(2)\,|\,Z(1)\}]. \tag{5.6}$$

Suppose that Σ_t is the forecast covariance matrix for $\mathbf{Y}_t$ such that

$$\{\Sigma_t\}_{jk} = \sigma_t(j, k) = \operatorname{cov}\{Y_t(j), Y_t(k)\,|\,\mathbf{y}^{t-1}\} \qquad j, k=1, \ldots, n$$

and let $\Sigma_t(r)$ be the forecast covariance matrix of $\{Y_t(1), \ldots, Y_t(r-1)\}$ for $r=2, \ldots, n$. Therefore by using identity (5.6) together with equations (5.4) and (5.5) it can be shown that

$$\begin{aligned} \sigma_t(1, 1) &= \tau_t^2(1), \\ \sigma_t(r, r) &= E[\operatorname{var}\{Y_t(r)\,|\,\mathbf{y}^{t-1}(r), \mathbf{x}^t(r)\}] + \operatorname{var}[E\{Y_t(r)\,|\,\mathbf{y}^{t-1}(r), \mathbf{x}^t(r)\}] \end{aligned}$$

$$= \tau_t^2(r) + \mathrm{trace}\{S_t(r)^{\mathrm{T}}\Sigma_t(r)S_t(r)\} + \bar{\mathbf{a}}_t(r)^{\mathrm{T}}\Sigma_t(r)\bar{\mathbf{a}}_t(r) \qquad 2 \leqslant r \leqslant n$$

where $\bar{\mathbf{a}}_t(r)^{\mathrm{T}} = \{\bar{a}_t^{(1)}(r), \ldots, \bar{a}_t^{(r-1)}(r)\}$.

To find the marginal covariance between $Y_t(r)$ and $Y_t(k)$, note that for the two variables $Z(1)$ and $Z(2)$

$$E\{Z(1)\,Z(2)\} = E[E\{Z(2)\,Z(1)\,|\,Z(2)\}] = E[Z(2)\,E\{Z(1)\,|\,Z(2)\}].$$

Therefore suppose that $Z(1) = Y_t(r)$ and $Z(2) = \mathbf{X}_t(r)$ and let $\boldsymbol{\sigma}_t(r)$ be a vector such that for $r \geqslant 2$

$$\boldsymbol{\sigma}_t(r)^{\mathrm{T}} = \{\sigma_t(r, 1), \ldots, \sigma_t(r, r-1)\} = \mathrm{cov}\{Y_t(r), \mathbf{X}_t(r)\,|\,\mathbf{y}^{t-1}\}.$$

Then

$$\boldsymbol{\sigma}_t(r) = E[\mathbf{X}_t(r)\,E\{Y_t(r)\,|\,\mathbf{X}_t(r)\}\,|\,\mathbf{y}^{t-1}].$$

By equation (5.4) this becomes

$$\boldsymbol{\sigma}_t(r) = \bar{a}_t^{(0)}(r)\,E\{\mathbf{X}_t(r)\,|\,\mathbf{y}^{t-1}\} + \sum_{i=1}^{r-1} \bar{a}_t^{(i)}(r)\,E[\mathbf{X}_t(r)\{Y_t(i) - \hat{\mathbf{f}}_t(i)\}\,|\,\mathbf{y}^{t-1}].$$

It is clear that

$$\Sigma_t(r+1) = \begin{pmatrix} \Sigma_t(r) & \boldsymbol{\sigma}_t(r) \\ \boldsymbol{\sigma}_t(r)^{\mathrm{T}} & \sigma_t(r, r) \end{pmatrix}.$$

So if $\Sigma_t(2) = \sigma(1, 1)$ is given, then $\Sigma_t(r+1)$ can be calculated as a simple function of $\Sigma_t(r)$, $\boldsymbol{\sigma}_t(r)$ and $\sigma_t(r, r)$. Hence the marginal forecast covariance matrix of the CLMDM is simply calculated from the variances and expectations of the updated distributions for the separate conditional component regression DLMs.

These models are best illustrated by a simple example.

5.1. *Example*

A simple example which illustrates this new class of models can be constructed as follows. Suppose that $\mathbf{Y}_t^{\mathrm{T}} = \{Y_t(1), Y_t(2)\}$ is to be modelled by an LMDM where $Y_t(1)$ represents the price of a certain brand relative to the rest of the market and $Y_t(2)$ represents the brand's market share. Suppose that at any fixed time the two variables can be represented by the graph of the influence diagram in Fig. 8(a). Suppose further that the processes $\{Y_k(1)\}_{k \leqslant t}$ and $\{Y_k(2)\}_{k \leqslant t}$ can be represented by the graph of the influence diagram of Fig. 8(b).

The general LMDM for this system is given by the following observation and system equations:

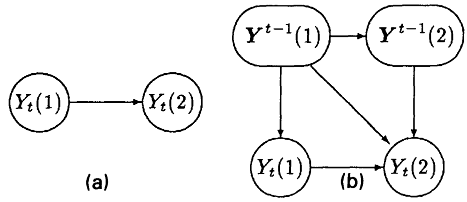

Fig. 8. Graph of the influence diagrams for (a) $Y_t(1)$ and $Y_t(2)$ for fixed time t and (b) $\{Y_k(1)\}_{k \leqslant t}$ and $\{Y_k(2)\}_{k \leqslant t}$

(a) *observation equation—*

$$Y_t(1) = \mathbf{F}_t(1)^{\mathrm{T}} \boldsymbol{\theta}_t(1) + v_t(1), \qquad v_t(1) \sim N(0, V_t(1)), \tag{5.7}$$

$$Y_t(2) = \mathbf{F}_t(2)^{\mathrm{T}} \boldsymbol{\theta}_t(2) + v_t(2), \qquad v_t(2) \sim N(0, V_t(2)); \tag{5.8}$$

(b) *system equation—*

$$\boldsymbol{\theta}_t = G_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \qquad \mathbf{w}_t \sim N(0, W_t); \tag{5.9}$$

(c) *initial information—*

$$(\boldsymbol{\theta}_0 | D_0) \sim N(\mathbf{m}_0, C_0) \tag{5.10}$$

where $\mathbf{F}_t(1) = \tilde{\mathbf{x}}_t(1)$, $\mathbf{F}_t(2)^{\mathrm{T}} = \{1, y_t(1)\}$, $\boldsymbol{\theta}_t^{\mathrm{T}} = \{\boldsymbol{\theta}_t(1)^{\mathrm{T}}, \theta_t^{(0)}(2), \theta_t^{(1)}(2)\}$, $\mathbf{m}_0^{\mathrm{T}} = \{\mathbf{m}_0(1)^{\mathrm{T}}, m_0^{(0)}(2), m_0^{(1)}(2)\}$, G_t is the s_1+2 square identity matrix (where s_1 is the dimension of $\boldsymbol{\theta}_t(1)\}$, W_t and C_0 are block diagonal and $v_t(1)$, $v_t(2)$, $\mathbf{w}_t(1)$ and $\mathbf{w}_t(2)$ are mutually independent of each other through time.

Using exactly the same notation introduced to define the means and variances of the LMDM, and noting that in this example $a_t^{(0)}(2) = m_{t-1}^{(0)}(2)$ and $a_t^{(1)}(2) = m_{t-1}^{(1)}(2)$, equations (5.2) and (5.5) lead directly to the required conditional forecast means and variances which are given by

$$E\{Y_t(2) | \mathbf{y}^{t-1}, \mathbf{y}_t(1)\} = a_t^{(0)}(2) + y_t(1)\, a_t^{(1)}(2) \tag{5.11}$$

and

$$\begin{aligned} \operatorname{var}\{Y_t(1) | \mathbf{y}^{t-1}(1)\} &= \tau_t^2(1), \\ \operatorname{var}\{Y_t(2) | \mathbf{y}^{t-1}, \mathbf{y}_t(1)\} &= y_t(1)^2 R_t^*(2) + \tau_t^2(2) \end{aligned} \tag{5.12}$$

where $\tau_t^2(2)$ is a function of $\tilde{R}_t(2)$, $R_t'(2)$ and $V_t(2)$ and in this case $R_t^*(2)$ is simply a scalar.

The first two moments of the forecast distribution are simply

$$E[\mathbf{Y}_t | \mathbf{y}^{t-1}] = \begin{pmatrix} a_t^{(0)}(1) \\ a_t^{(0)}(2) + a_t^{(0)}(1)\, a_t^{(1)}(2) \end{pmatrix} \tag{5.13}$$

and covariance matrix

$$\Sigma = \begin{pmatrix} \tau_t^2(1) & \sigma_t(1,2) \\ \sigma_t(2,1) & \sigma_t(2,2) \end{pmatrix} \tag{5.14}$$

when

$$\sigma_t(1,2) = \sigma_t(2,1) = a_t^{(0)}(2)\, E\{Y_t(1) | \mathbf{y}^{t-1}\} + a_t^{(1)}(2)\, E\{Y_t(1)^2 | \mathbf{y}^{t-1}\}$$

and

$$\sigma_t(2,2) = \tau_t^2(2) + R_t^*(2)\, E\{Y_t(1)^2 | \mathbf{y}^{t-1}\} + a_t^{(1)}(2)^2\, \tau_t^2(1).$$

A time series plot of $\{Y_k(2)\}_{k>0}$ is given in Fig. 9 together with various one-step-ahead forecasts of the series. The time series itself is denoted by the dots on the graph. The one-step-ahead conditional and marginal forecasts of $Y(2)$ derived from the MDM are represented by the thin and thick curves respectively. The conditional forecasts obtained from the MDM are fairly accurate, indicating that the relative price of the brand appears to be a good predictor of the brand's market share. The one-step-

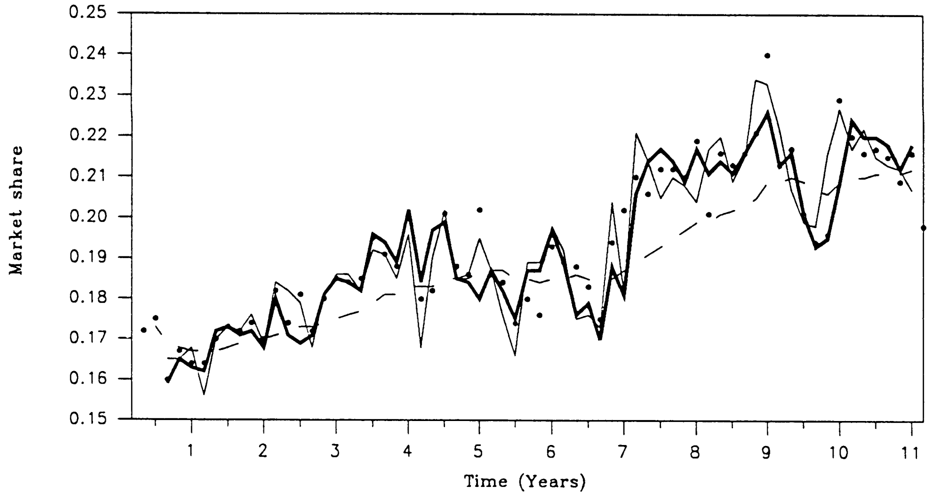

Fig. 9. Time series plot of $Y(2)$ with one-step-ahead forecasts: ······, time series; ———, one-step-ahead forecast conditional on relative price; ▬▬▬, one-step-ahead marginal forecast derived from the MDM; – – – –, forecasts from the steady model

ahead forecasts of the simple steady model for $Y(2)$ are represented by the broken curve in Fig. 9. If these are compared directly with the marginal forecasts of the MDM, then clearly the forecasts obtained from the MDM are far superior to those obtained from the steady model. This illustrates how the forecast performance of a model for $Y(2)$ can be improved by using $Y(1)$ as a regressor and even though $Y_t(1)$ cannot be observed before $Y_t(2)$ the MDM can still use $Y_t(1)$ to obtain realistic forecasts of $Y_t(2)$.

Clearly, both $\{Y_t(1)|\mathbf{y}^{t-1}(1)\}$ and $\{Y_t(2)|\mathbf{y}^{t-1}, y_t(1)\}$ have normal distributions. However, the joint forecast distribution of $\{Y_t(1), Y_t(2)|\mathbf{y}^{t-1}\}$ is *not* bivariate normal because of the appearance of $y_t(1)$ in the variance term of $\{Y_t(2)|\mathbf{y}^{t-1}, y_t(1)\}$. Indeed this joint distribution can be very non-normal. This is demonstrated in Fig. 10 in which the contours of the joint distribution of $\{Y_t(1), Y_t(2)\}$ after the margin of $Y_t(1)$ has been normalized and for various parameter values are given.

From Fig. 10 it can be shown that the modes and antimodes lie on a quintic (and so exhibit a butterfly catastrophe; Zeeman (1977)). The joint forecast distribution is only symmetrical when the variance of $\{Y_t(1)|\mathbf{y}^{t-1}\}$ does not appear in the variance of $\{Y_t(2)|\mathbf{y}^{t-1}\}$, i.e. when $m_{t-1}^{(1)}(2)=0$. For any non-degenerate values of the parameters, it can be concluded, after a little algebra, that there is a value of $y_t(2)$ such that the conditional predictive density of $\{Y_t(1)|\, Y_t(2)=y_t(2)\}$ is bimodal. The worst case occurs when $\boldsymbol{\theta}_t(2)$ is uncertain where the distribution becomes very non-Gaussian. As $R_t^*(2) \to 0$, then the process tends to a bivariate normal distribution. However, unlike the analogous simultaneous equations models the assumed stochastic drift on $\boldsymbol{\theta}_t(2)$ prevents this limit from ever being reached.

From Fig. 10 it is clear that the point $E[Y_t(1)|\mathbf{y}^{t-1}(1)]$ (in this case $Y_t(1)=0$) takes on special significance as it is the point about which the contours are symmetrical or asymmetrical. Regression on the forecast residual $[Y_t(1) - E\{Y_t(1)|\mathbf{y}^{t-1}(1)\}]$ will often therefore seem more natural. This gives the corresponding CLMDM whose means and covariance matrix are given by

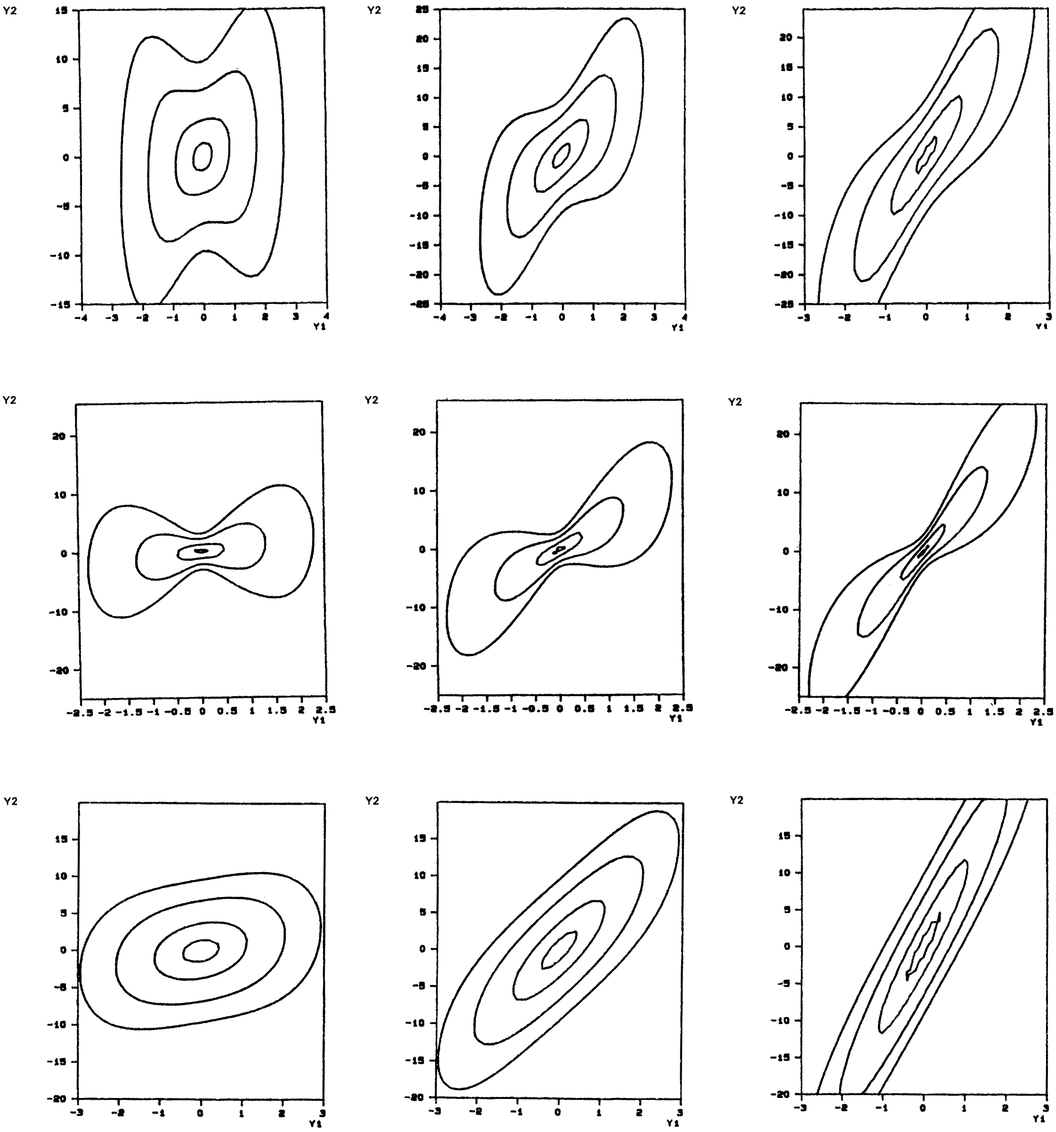

Fig. 10. Various asymmetric contour plots of $Y_t(1)$ and $Y_t(2)$ when the marginal distribution of $Y_t(1)$ has been normalized to have 0 mean and unit variance and $m_{t-1}(2) \neq 0$; $R_t^*(2) = \tau_t^2(2) = 10$ in the first row, $R_t^*(2) = 10$, $\tau_t^2(2) = 1$ in the second row and $R_t^*(2) = 1$, $\tau_t^2(2) = 10$ in the third row; $m_{t-1}(2)$ varies by column, taking the values 1 in the first column, 5 in the second and 10 in the third

$$E[\mathbf{Y}_t|\mathbf{y}^{t-1}] = \begin{pmatrix} a_t^{(0)}(1) \\ 0 \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} \tau_t^2(1) & \sigma_t(1, 2) \\ \sigma_t(2, 1) & \sigma_t(2, 2) \end{pmatrix}$$

where

$$\sigma_t(1,2) = \sigma_t(2,1) = a_t^{(1)}(2)\,E[\{Y_t(1)-\hat{\mathbf{f}}_t(1)\}^2\,|\,\mathbf{y}^{t-1}],$$

$$\begin{aligned}\sigma_t(2,2) &= \tau_t^2(2) + R_t^*(2)\,E[\{Y_t(1)-\hat{\mathbf{f}}_t(1)\}^2\,|\,\mathbf{y}^{t-1}] + a_t^{(1)}(2)^2\,\mathrm{var}[\{Y_t(1)-\hat{\mathbf{f}}_t(1)\}\,|\,\mathbf{y}^{t-1}]\\ &= \tau_t^2(2) + \tau_t^2(1)\{R_t^*(2)+a_t^{(1)}(2)^2\}.\end{aligned}$$

The graphs in Fig. 10 will still be appropriate for this corrected model.

6. DISCUSSION OF MULTIREGRESSION DYNAMIC MODELS

Some of the interesting theoretical aspects of MDMs will be discussed in this section.

MDMs define a class of non-Gaussian time series models which decompose the forecasting system into components whose conditional distributions are univariate Bayesian dynamic regression models. As was mentioned in Sections 3 and 5, these univariate models can be normal. In particular, if $\mathbf{F}_t(r)$ is a linear function of $\{\mathbf{x}^t(r), \mathbf{y}^{t-1}(r)\}$, then each variable would follow a generalization of one of Priestley's state-dependent models (Priestley, 1980). Alternatively, each model could be a non-linear dynamic model (West *et al.*, 1985; Migon and Harrison, 1985; Pole *et al.*, 1988). If each component follows a univariate Bayesian linear or non-linear model, then it is possible to use the updating relationships directly from univariate Bayesian dynamic models, even though the model regresses on contemporary components of $\mathbf{Y}_t$ and can be highly non-linear. This make the process especially interesting because the models are amenable to analytical investigation. Approximate or numerical methods, with all the robustness issues that surround them, are largely unnecesary. However, as can be seen from Fig. 10, the vector $\mathbf{Y}_t$ can have a joint forecast distribution which is very non-Gaussian, even when $\mathbf{F}_t$ is a linear function of $(Y_t(1), \ldots, Y_t(r-1))$. In this respect this class of models is very similar to the models of Wermuth and Lauritzen (1990) in the fact that the conditional distributions are fairly simple but the joint model is far more complex.

It is interesting that, unlike conditional independence within time frames, conditional causality of the type described here depends on the ordering of the variables in the MDM. Furthermore each influence diagram corresponds to a unique LMDM or CLMDM structure.

For example, suppose that for five variables modelled with an LMDM the following regression equations hold:

$$\left.\begin{aligned}Y_t(1) &= \theta_t^{(0)}(1) + v_t(1);\\ Y_t(2) &= \theta_t^{(0)}(2) + \theta_t^{(1)}(2)\,y_t(1) + v_t(2);\\ Y_t(3) &= \theta_t^{(0)}(3) + \theta_t^{(1)}(3)\,y_t(1) + \theta_t^{(2)}(3)\,y_t(2) + v_t(3);\\ Y_t(4) &= \theta_t^{(0)}(4) + \theta_t^{(3)}(4)\,y_t(3) + v_t(4);\\ Y_t(5) &= \theta_t^{(0)}(5) + \theta_t^{(3)}(5)\,y_t(3) + v_t(5).\end{aligned}\right\} \tag{6.1}$$

The graph of the influence diagram consistent with these equations is given by graph (a) in Fig. 11.

It is easy to check that the LMDM on these regression equations could alternatively be represented by the LMDM on components of $\mathbf{Y}$ taken in the order $\{Y(1), Y(2),$

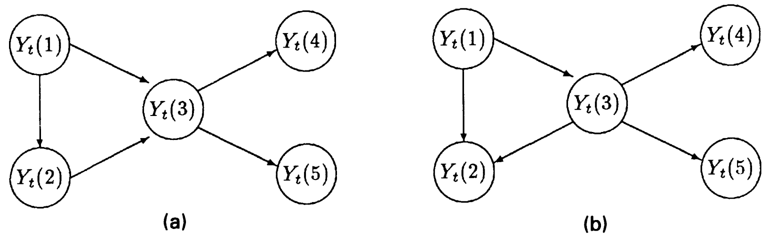

Fig. 11. Two graphs with the same implied conditional independences, but with different LMDMs

$Y(3)$, $Y(5)$, $Y(4)\}$. The graph of the influence diagram, Fig. 11(a), would remain unchanged.

In general any new ordering of the variables that is compatible with the influence diagram of a single time frame of an LMDM or CLMDM, given the past, will produce equations that are algebraically identical with the original.

However by the decomposition theorem (Smith, 1989), for any given time frame t, the implied set of conditional independence statements by graph (b) in Fig. 11, given the past, are identical with those represented by graph (a), but the conditional independences *related to causality* in the two diagrams correspond to two quite different LMDMs.

For example, from graph (a) it is clear that

$$Y_t(2)\,|\,\mathbf{y}^{t-1} = Y_t(2)\,|\,\mathbf{y}^{t-1}(2), \mathbf{y}^{t-1}(1)$$

whereas graph (b) means that

$$Y_t(2)\,|\,\mathbf{y}^{t-1} = Y_t(2)\,|\,\mathbf{y}^{t-1}(2), \mathbf{y}^{t-1}(1), \mathbf{y}^{t-1}(3).$$

Because of the different covariance structure on the system error $\mathbf{w}_t$ implied by each of these influence diagrams, there is no guarantee that these two statements could hold simultaneously.

Suppose that the context of the model is such that the time frame conditional independences are logically determined. However, suppose that there is uncertainty regarding the causal structure across the variables in the problem, although this causal structure is assumed consistent over time. In this cases a class I *multiprocess MDM* (Harrison and Stevens, 1976) can model the system. A multiprocess MDM considers m models $\{M^{(1)}, \ldots, M^{(m)}\}$ simultaneously where there is one model for each of the m possible causal structures for the given conditional independence structure. They allow the prediction of complex series without any *a priori* assumptions about causal structures. This is because the forecasts are found by using $p(\mathbf{y}_t\,|\,\mathbf{y}^{t-1}, M^{(i)})$ mixed with probabilities $p(M^{(i)}\,|\,\mathbf{y}^{t-1})$ to give appropriate predictive densities. Multiprocess models can also be combined with some decision rule which compares the forecast performance of the various models so that the most appropriate causal structure can be selected from the n alternatives.

Although Granger causality has attracted academic interest, in practice it has proved difficult to discriminate between different causal structures by using linear systems. However, by embedding causality in the non-linear MDM, multiprocess MDMs can give an *on-line* assessment of hypothetical causal relationships across the variables. The MDM corresponding to a given causal structure can then be made at least plausible. Furthermore, the conditional components can be as complicated as necessary, containing trends, regression terms, seasonal factors, and so on. It therefore looks promising that the MDMs will enable the selection of causal structures across practical models.

Notice that from the geometries of the joint densities on the model of Section 5.1 most information about causal relationships seems to come when the relationship between the variables is uncertain (in this example this corresponds to $R_t^*(2)$ being large). This will occur early in the series and after external intervention. This might explain why Zellner (1984) finds it so difficult to discriminate between two causal structures—the models that he considers are not dynamic; they would assume that $R_t^*(2) \rightarrow 0$ and external intervention is not considered.

7. CONCLUSION

MDMs have been developed to model multivariate time series whose components exhibit conditional independences related to causality. Although highly non-linear, the models are relatively simple to implement. Preliminary use of these models on series with a small number of components seems promising. Use of the MDM to model a large business market is documented in Queen (1992).

ACKNOWLEDGEMENTS

We would like to thank the Science and Engineering Research Council and Unilever Research for the financial support given while undertaking this research.

REFERENCES

Dawid, A. P. (1979) Conditional independence in statistical theory (with discussion). *J. R. Statist. Soc.* B, **41**, 1–31.

Florens, J. P. and Mouchart, M. (1985) A linear theory for non-causality. *Econometrica*, **53**, 157–175.

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

Harrison, P. J. and Stevens, C. F. (1976) Bayesian forecasting (with discussion). *J. R. Statist. Soc.* B, **38**, 205–217.

Harvey, A. C. (1986) Analysis and generalisation of a multivariate exponential smoothing model. *Mangmnt Sci.*, **32**, 374–380.

——— (1989) *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Holland, P. W. (1986) Statistics and causal inference. *J. Am. Statist. Ass.*, **81**, 945–970.

Howard, R. A. and Matheson, J. E. (1981) Influence diagrams. In *Readings on the Principles and Applications of Decision Analysis* (eds R. A. Howard and J. E. Matheson), vol. II, pp. 719–762. Menlo Park: Strategic Decision Group.

Kitagawa, G. (1987) Non-Gaussian state-space modelling of nonstationary time series (with discussion). *J. Am. Statist. Ass.*, **82**, 1032–1063.

Lauritzen, S. L. (1989) Mixed graphical association models. *Scand. J. Statist.*, **16**, 273–306.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N. and Leimer, H. G. (1990) Independence properties of directed Markov fields. *Networks*, **20**, 491–505.

Migon, H. S. and Harrison, P. J. (1985) An application of non-linear Bayesian forecasting to television and advertising. In *Bayesian Statistics 2* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith). Amsterdam: North-Holland.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo: Morgan Kaufmann.

Pearl, J. and Verma, T. S. (1987) The logic of representing dependencies by directed graphs. In *Proc. 6th Natn. Conf. Artificial Intelligence, Seattle*, pp. 374–379. Los Altos: Morgan Kaufmann.
Pole, A. and West, M. (1990) Efficient Bayesian learning in non-linear dynamic models. *J. Forecast.*, **9**, 119–136.
Pole, A., West, M. and Harrison, P. J. (1988) Non-normal and non-linear dynamic Bayesian modelling. In *Bayesian Analysis of Time Series and Dynamic Models* (ed. J. C. Spall). New York: Dekker.
Priestley, M. B. (1980) State-dependent models: a general approach to non-linear time series analysis. *J. Time Ser. Anal.*, **1**, 47–71.
Queen, C. M. (1992) Using the multiregression dynamic model to forecast the brand sales in a competitive product market. *Research Report 92-10.* Nottingham Statistics Group, University of Nottingham, Nottingham.
Quintana, J. M. and West, M. (1987) Multivariate time series analysis: new techniques applied to international exchange rate data. *Statistician*, **36**, 275–281.
Shachter, R. D. (1986) Intelligent probabilistic inference. In *Uncertainty and Artificial Intelligence* (eds L. N. Kanal and J. Lemmer), pp. 371–382. Amsterdam: North-Holland.
Smith, J. Q. (1989) Influence diagrams for statistical modelling. *Ann. Statist.*, **17**, 654–672.
——— (1990) Statistical principles of graphs. In *Influence Diagrams, Belief Nets and Decision Analysis* (eds R. M. Oliver and J. Q. Smith), pp. 89–120. New York: Wiley.
Wermuth, N. and Lauritzen, S. L. (1990) On substantive research hypotheses, conditional independence graphs and graphical chain models. *J. R. Statist. Soc.* B, **52**, 21–50.
West, M. and Harrison, P. J. (1989) *Bayesian Forecasting and Dynamic Models*. New York: Springer.
West, M., Harrison, P. J. and Migon, H. J. (1985) Dynamic generalised linear models and Bayesian forecasting (with discussion). *J. Am. Statist. Ass.*, **80**, 73–97.
Zeeman, E. C. (1977) *Catastrophe Theory: Selected Papers (1972–1977)*. Reading: Addison-Wesley.
Zellner, A. (1984) *Basic Issues in Econometrics*. Chicago: University of Chicago Press.