Intervention and causality in a dynamic Bayesian network

Catriona M. Queen[1] **and Casper J. Albers**

The Open University, UK

January 10, 2008

Abstract

The use of intervention for time series modelling is a well established technique for on-line forecasting and decision-making in the context of Bayesian dynamic linear models. Intervention has also been recently used in (non-dynamic) Bayesian networks to investigate causal relationships between variables, and in dynamic Bayesian networks to investigate lagged causal relationships between time series. The Multiregression Dynamic Model (MDM) is a Bayesian dynamic model and an example of a dynamic Bayesian network. The focus of this paper is the use of intervention in the MDM. It will be demonstrated that not only is intervention in the MDM a powerful tool for forecasting, but intervention can also aid in the identification of contemporaneous causal relationships between time series, thus going beyond the identification of lagged causal relationships previously addressed in dynamic Bayesian networks.

Keywords: Intervention, causality, Bayesian forecasting, dynamic linear model, multiregression dynamic model

1 Introduction

A Bayesian network (BN) is a directed acyclic graph in which variables are represented by nodes and arcs between nodes represent conditional dependencies between the variables. A dynamic Bayesian network (DBN) is a Bayesian network for a sequence of variables such as a time series or stochastic process. DBNs of various forms have received a lot of interest in recent years (see for example Brillinger (1996), Dahlhaus (2000), Dahlhaus and Eichler (2003), Farrow *et al.* (1997), Kjærulff (1995), Smith and Figueroa (2007), Sun *et al.* (2006)).

[1]Contact details: Department of Statistics, The Open University, Milton Keynes, MK7 6AA, UK. E-mail: C.Queen@open.ac.uk, Tel: (+44) 01908 659585, Fax:(+44) 01908 655515

The multiregression dynamic model (MDM) (Queen and Smith, 1993) is an example of a DBN. The MDM is a Bayesian dynamic model and is defined to preserve any conditional independences, related to causality, across a multivariate time series over time. At each time t, the observable component series $Y_t(1), \ldots, Y_t(n)$ of the n-dimensional time series, and their associated state vectors $\boldsymbol{\theta}_t(1), \ldots, \boldsymbol{\theta}_t(n)$, are represented by a BN. These individual BNs are linked together over time to form a DBN. The MDM then uses the conditional independences and causal driving mechanism through the system, as represented by the DBN, to break down the multivariate model into simpler univariate components.

There are many potential application areas for the MDM, including problems in economics (modelling various economic indicators such as energy consumption and GDP), environmental problems (such as water and other resource management problems), industrial problems (such as product distribution flow problems) and medical problems (such as patient physiological monitoring). Queen (1994) and Queen *et al.* (2007b) use an MDM to model monthly brand sales in a competitive market. Here, the competition in the market is the causal drive within the system and is used to define a conditional independence structure across the time series. Queen (1997) and Queen *et al.* (1994) focus on how the DBN at time t may be elicited for MDMs for modelling markets. Following Whitlock and Queen (2000) and Queen *et al.* (2007a), this paper considers (in Section 4) the specific application of forecasting traffic flows at various points in a road network. In this application, the direction of traffic flow produces the causal drive through the system and the possible routes through the system are used to define a conditional independence structure across the time series.

When using the MDM for on-line forecasting and decision-making, it is essential that the model is able to accommodate external intervention, or manipulation, of the system. For example, when using an MDM to forecast traffic flows in a road network, various factors external to the system, such as roadworks or bad weather, can affect the traffic flows in the network. If the forecast performance of the MDM is to be maintained, then the model needs to be manipulated to take account of this information.

The effects of introducing external interventions into BNs in a non-dynamic setting have been addressed through causal Bayesian networks (CBNs) (Pearl, 1995, 2000; Spirtes *et al*, 2000; Lauritzen, 2000; Lauritzen and Richardson, 2002; Dawid, 2002). Here interventions are used to identify causal relationships between the variables. More explicitly, suppose that a set of variables $X_1, \ldots, X_n$ are represented by a BN. If there is a directed arc from X_i to X_j in the BN, then X_i is said to be a parent of X_j. Then the joint density $f(\boldsymbol{x})$ is given by the factorisation

$$f(\boldsymbol{x}) = \prod_{j=1}^{n} f(x_j|pa_j),$$

where pa_j denotes the set of variables in $X_1, \ldots, X_n$ which are parents of X_j in the BN. Let $\boldsymbol{X}_{-i}$ denote the variables $\{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$. Then, conditional on observing the value $X_i = x_i$, the density for the rest of the variables is given by

$$f(\boldsymbol{x}_{-i}|x_i) = \frac{f(\boldsymbol{x})}{f(x_i)}.$$

Now suppose that the value of X_i is manipulated externally, instead of arising naturally by observation, and is assigned the value x_i. For example, X_i could be the prescribed dose of a drug given to a patient. In this case, the density for the rest of the variables, conditional on the *intervention* $X_i = x_i$, denoted $f(\boldsymbol{x}_{-i}\|x_i)$ (following Lauritzen (2000)), is generally *not* the same as $f(\boldsymbol{x}_{-i}|x_i)$, so that

$$f(\boldsymbol{x}_{-i}\|x_i) = \prod_{j=1, j\neq i}^{n} f(x_j|pa_j(x_i)) = \frac{f(\boldsymbol{x})}{f(x_i|pa_i)},$$

where $pa_j(x_i)$ denotes the parents of x_j with X_i set to the value x_i. In this case, the Bayesian network is a CBN. Essentially, in a CBN, X_i is causal for X_j if intervention with respect to X_i affects X_j's distribution. Further, an intervention with respect to X_i can only affect its descendants so that X_i is causal for its descendants.

Didelez (2003a) and Eichler and Didelez (2007) consider the effects of external intervention in a dynamic setting for DBNs. Here, a time series $\{X_t(i)\}$ is said to be causal for the time series $\{X_t(j)\}$ if an intervention with respect to $X_t(i)$ affects the predictions of $X_{t+k}(j)$, for some future time $t + k$. Thus intervention in DBNs has been used to investigate lagged causal relationships between time series.

The focus of this paper is the use of intervention in the MDM. It will be demonstrated that not only is intervention in the MDM a powerful tool for forecasting, but it can also aid the identification of contemporaneous causal relationships between time series, thus going beyond the identification of lagged causal relationships previously addressed in DBNs.

Following Dawid (2002), this paper uses influence diagrams to investigate the effects of interventions. An influence diagram (ID) is a generalisation of a BN which can be used to represent and solve Bayesian decision problems (Howard and Matheson, 1984; Shachter, 1986, 1988; Oliver and Smith, 1990). More explicitly, an ID is a directed acyclic graph with random nodes, drawn as ovals, representing the variables, and decision nodes, drawn as rectangles, representing decisions. The value of a decision node arises through external intervention by a decision maker, and is not assumed to arise naturally through observation. As in a BN, arcs leading into random nodes represent conditional dependencies, so that each random node has an associated distribution, conditional on its parents, just as in a BN. Arcs leading into decision nodes represent information flow, so that information regarding its parents is assumed available to the decision maker before a decision is made. An ID can be used in exactly the same way as a BN to investigate conditional independences between the variables, both random and decision.

Throughout the paper, the standard notation of Dawid (1979) for conditional independence will be used, so that $A \perp\!\!\!\perp B|C$ reads "A is independent of B given C". The paper is structured as follows. Section 2 provides an introduction to the MDM before focussing on intervention in the MDM in Section 3. In Section 4, intervention in the MDM is applied to the problem of forecasting hourly traffic flows at a junction of three major roads in the UK. Section 5 shows how the MDM can be used to identify contemporaneous causal relationships between series. Finally, Section 6 offers some concluding remarks.

2 The Multiregression Dynamic Model

In this section, the MDM of Queen and Smith (1993) will be defined.

Let $\boldsymbol{Y}_t = (Y_t(1), \ldots, Y_t(n))^\top$ denote the n-dimensional multivariate time series and let $\boldsymbol{Y}^t = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t)^\top$ and $\boldsymbol{Y}^t(i) = (Y_1(i), \ldots, Y_t(i))^\top$. Suppose that there is a causal driving mechanism across the series and that the series is ordered so that the same conditional independence structure related to causality is defined across the series through time, where at each time $t \in \mathbb{N}$,

$$Y_t(i) \amalg \{\{Y_t(1), \ldots, Y_t(i-1)\} \setminus pa(Y_t(i))\} \, | pa(Y_t(i)), \qquad \text{for } i = 2, \ldots n.$$

The notation "\" reads "excluding" and $pa(Y_t(i)) \subseteq \{Y_t(1), \ldots, Y_t(i-1)\}$. Each variable in the set $pa(Y_t(i))$ is a parent of $Y_t(i)$. The conditional independence relationships at each time point t can be represented by a BN, where there is a directed arc to $Y_t(i)$ from each of its parents in $pa(Y_t(i))$. To illustrate, Figure 1 shows a BN for five time series at time t, where $pa(Y_t(2)) = \emptyset$, $pa(Y_t(3)) = \{Y_t(1), Y_t(2)\}$, $pa(Y_t(4)) = \emptyset$ and $pa(Y_t(5)) = \{Y_t(2), Y_t(3), Y_t(4)\}$.

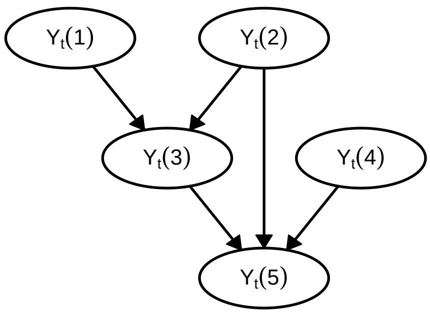

Figure 1: Bayesian network representing five time series at time t.

Suppose further that there is a conditional independence structure related to causality defined for the processes over time so that,

$$Y_t(i) \amalg \big\{ \big\{ \boldsymbol{Y}^t(1), \ldots, \boldsymbol{Y}^t(i-1) \big\} \setminus pa(\boldsymbol{Y}^t(i)) \big\} \, | (pa(\boldsymbol{Y}^t(i)), \boldsymbol{Y}^{t-1}(i)), \qquad \text{for } i = 2, \ldots n.$$

Thus the conditional independence relationships over time are represented by a DBN.

Denote the information available at time t by D_t. The MDM for the n-dimensional vector time series $\boldsymbol{Y}_t$ over times $t = 1, 2, ...,$ is defined by the n observation equations, the system equation, and information at time $t-1$ as follows.

$$\begin{array}{llcll}
\text{Observation equations:} & Y_t(i) & = & \boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + v_t(i), & v_t(i) \sim (0, V_t(i)) \\
 & & & & 1 \leq i \leq n \\
\text{System equation:} & \boldsymbol{\theta}_t & = & G_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, & \boldsymbol{w}_t \sim (\boldsymbol{0}, W_t) \\
\text{Information:} & (\boldsymbol{\theta}_{t-1} \,|\, D_{t-1}) & \sim & (\boldsymbol{m}_{t-1}, C_{t-1}). &
\end{array}$$

The m_i-dimensional vector $\boldsymbol{F}_t(i)$ contains an arbitrary, but known, function of the parents $pa(Y_t(i))$ and possibly other known variables (which may include $\boldsymbol{Y}^{t-1}$); $\boldsymbol{\theta}_t(i)$ is the m_i-dimensional parameter vector for $Y_t(i)$; $V_t(1), \ldots V_t(n)$ are the scalar observation variances; $\boldsymbol{\theta}_t^\top = (\boldsymbol{\theta}_t(1)^\top, \ldots, \boldsymbol{\theta}_t(n)^\top)$ is the m-dimensional parameter vector; $\boldsymbol{m}_{t-1}$ and C_{t-1} are the (posterior) moments for $\boldsymbol{\theta}_{t-1}$ at time $t-1$; and the $m \times m$ matrices $G_t = \text{blockdiag}\{G_t(1), \ldots, G_t(n)\}$, $W_t = \text{blockdiag}\{W_t(1), \ldots, W_t(n)\}$ and $C_{t-1} = \text{blockdiag}\{C_{t-1}(1), \ldots, C_{t-1}(n)\}$ are such that $G_t(i)$, $W_t(i)$ and $C_{t-1}(i)$ are $m_i \times m_i$ square matrices assumed known (and, in particular, are not functions of $pa(Y_t(i))$). The error vectors, $\boldsymbol{v}_t^\top = (v_t(1), \ldots, v_t(n))$ and $\boldsymbol{w}_t^\top = (\boldsymbol{w}_t(1)^\top, \ldots, \boldsymbol{w}_t(n)^\top)$, are such that $v_t(1), \ldots, v_t(n)$ and $\boldsymbol{w}_t(1), \ldots, \boldsymbol{w}_t(n)$ are mutually independent and $\{\boldsymbol{v}_t, \boldsymbol{w}_t\}_{t \in \mathbb{N}}$ are mutually independent with time.

Note that no distributional assumptions have been placed on the error terms or the distribution for $\boldsymbol{\theta}_{t-1}$. Also, there is no specific requirement that $\boldsymbol{F}_t(i)$ be a linear function of $pa(Y_t(i))$, just that the function is known. Thus the MDM is a very general model. Each $Y_t(i)|(pa(Y_t(i)), \boldsymbol{\theta}_t(i))$ therefore follows some distribution with mean $\boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i)$ and variance $V_t(i)$. Similarly, $\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}$ follows some distribution with mean $G_t \boldsymbol{\theta}_{t-1}$ and covariance W_t.

The MDM uses the conditional independence structure to model the multivariate time series by n separate univariate models — for $Y_t(1)$ and $Y_t(i)|pa(Y_t(i))$, $i = 2, \ldots, n$. Each $Y_t(i)$ is then modelled by a Bayesian univariate dynamic regression model (see West and Harrison (1997) for details) with some function of its parents as regressors. For each $Y_t(i)$ without parents, any suitable Bayesian univariate dynamic model is used. For example, consider the BN in Figure 1. Neither $Y_t(1)$, $Y_t(2)$ nor $Y_t(4)$ have parents and so each of these series can be modelled separately in an MDM using any suitable univariate dynamic models. The variables $Y_t(3)$ and $Y_t(5)$ both have parents and so these would be modelled by (separate) univariate dynamic regression models with functions of the two regressors $Y_t(1)$ and $Y_t(2)$ for $Y_t(3)$'s model, and functions of the three regressors $Y_t(2)$, $Y_t(3)$ and $Y_t(4)$ for $Y_t(5)$'s model.

As long as $\boldsymbol{\theta}_t(1), \boldsymbol{\theta}_t(2), \ldots, \boldsymbol{\theta}_t(n)$ are mutually independent initially (i.e. C_0 is block diagonal), then the block diagonal form of G_t and W_t ensure that each $\boldsymbol{\theta}_t(i)$ can be updated separately in closed form from $Y_t(i)$'s (conditional) univariate model. A forecast for $Y_t(1)$ and the conditional forecasts for $Y_t(i)|pa(Y_t(i))$, $i = 2, \ldots, n$, can also then be found separately, often using established dynamic model results. For example, in the context of the BN in Figure 1, forecasts can be found separately for

$$Y_t(1), \quad Y_t(2), \quad Y_t(3)|Y_t(1), Y_t(2), \quad Y_t(4) \quad \text{and} \quad Y_t(5)|Y_t(2)Y_t(3), Y_t(4).$$

However, $Y_t(i)$ and $pa(Y_t(i))$ are observed simultaneously. So the marginal forecast for each $Y_t(i)$, without conditioning on the values of its parents, is required. Unfortunately, the marginal forecast distributions for $Y_t(i)$, $i = 2, \ldots, n$, will not generally be of a simple form. However, (under quadratic loss) the moments of the marginal forecast distributions for $Y_t(i)$, $i = 2, \ldots, n$, are adequate for forecasting purposes, and these can be easily found for many MDMs.

When $\boldsymbol{F}_t(i)$ is a linear function of $pa(Y_t(i))$ and the error distributions are normal, then this is the special case of the Linear Multiregression Dynamic Model (LMDM). In this case, each $Y_t(i)$ with parents is modelled by a regression dynamic linear model (DLM) (Harrison and Stevens, 1976) with its parents as (linear) regressors, and each $Y_t(i)$ without parents is modelled by any appropriate DLM. LMDMs are particularly simple to use analytically and their use has been demonstrated for modelling monthly brand sales in competitive markets (Queen, 1994; Queen *et al.*, 2007b) and for modelling hourly vehicle counts at various points in a traffic network (Whitlock and Queen, 2000; Queen *et al.*, 2007a). An LMDM will be used to forecast traffic flows in a traffic network in Section 4.

It is important to realise that although two BNs can represent the same conditional independences, they can represent quite different conditional independences related to causality and consequently, quite different MDMs. For example, consider the two BNs in Figure 2. For both of these BNs, $Y_t(1) \perp\!\!\!\perp Y_t(3)|Y_t(2)$, however, the conditional independences related to causality relate to two quite different MDMs. For BN (a), an MDM would model $Y_t(1)$, $Y_t(2)|Y_t(1)$ and $Y_t(3)|Y_t(2)$, whereas for BN (b), an MDM would model $Y_t(1)|Y_t(2)$, $Y_t(2)|Y_t(3)$ and $Y_t(3)$.

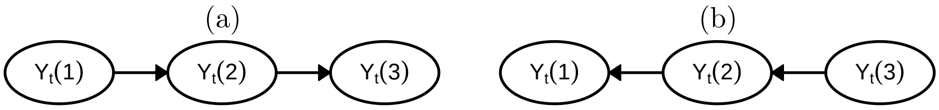

Figure 2: Two Bayesian networks representing the same conditional independences, but quite different MDMs.

3 Intervention in the Multiregression Dynamic Model

The accommodation of external intervention is essential for any model for on-line forecasting and decision-making. The use of interventions for time series modelling is a well established technique in the context of the DLM (Harrison and Stevens, 1976; West and Harrison, 1986, 1989, 1997). The intervention described in the introduction, in which X_i was assigned a specific (single) value x_i, is referred to as an 'atomic' intervention. The interventions used in DLMs are not atomic. Instead they are what are referred to as 'random' interventions in which the *distribution* of the random variable is manipulated. At each time t, a decision is made as to whether to use intervention for the (vector) series $\boldsymbol{Y}_t$ or not. Thus, sequential decisions are made at each time point regarding intervention on the time series, as in Didelez (2003b). In addition, interventions for the state vector are also permitted in the DLM, so that a further set of sequential decisions are made at each time point regarding intervention. As is assumed in Lauritzen (2000) and Dawid (2002), intervention in the DLM context always precedes observation, so that intervention at time t is done before forecasts are made and the series is observed at time t.

As with the DLM, random intervention is possible in the MDM for the time series $\boldsymbol{Y}_t$ and the parameter vector $\boldsymbol{\theta}_t$ at each time t. In addition, it is possible to intervene (separately) for any number and combination of the individual series $Y_t(1), \dots, Y_t(n)$ and/or their associated parameter vectors $\boldsymbol{\theta}_t(1), \dots, \boldsymbol{\theta}_t(n)$ at any particular time t. For clarity here, only the interventions at the single time point t for a single component $Y_t(i)$, and its associated parameter vector $\boldsymbol{\theta}_t(i)$, will be considered.

Explicitly, intervention in the MDM usually works as follows. The observation equation for $Y_t(i)$ specifies the distribution $Y_t(i)|(\boldsymbol{\theta}_t(i), pa(Y_t(i))) \sim (\boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i), V_t(i))$. Intervention for $Y_t(i)$ in the MDM therefore involves manipulating $Y_t(i)$'s observation equation so that

$$Y_t(i) = \boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + \boldsymbol{h}_t(i) + \boldsymbol{v}_t(i), \qquad \boldsymbol{v}_t(i) \sim (0, V_t(i) + H_t(i))$$

for some vector $\boldsymbol{h}_t(i)$ and matrix $H_t(i)$, which yields the intervention distribution

$$Y_t(i)|(\boldsymbol{\theta}_t(i), pa(Y_t(i))) \sim (\boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + \boldsymbol{h}_t(i), V_t(i) + H_t(i)).$$

The system equation specifies the distribution $\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1} \sim (G_t\boldsymbol{\theta}_{t-1}, W_t)$. So when intervening for $\boldsymbol{\theta}_t(i)$, the part of the system equation associated with $\boldsymbol{\theta}_t(i)$ is manipulated so that

$$\boldsymbol{\theta}_t(i) = G_t(i)^* \boldsymbol{\theta}_{t-1}(i) + \boldsymbol{w}_t(i), \qquad \boldsymbol{w}_t(i) \sim (\boldsymbol{0}, W_t(i)^*)$$

for suitable matrices $G_t(i)^*$ and $W_t(i)^*$, which yields the intervention distribution

$$\boldsymbol{\theta}_t(i)|\boldsymbol{\theta}_{t-1}(i) \sim (G_t(i)^* \boldsymbol{\theta}_{t-1}(i), W_t(i)^*).$$

Following the ideas of Dawid (2002), introduce indicator variables $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$, where

$$\begin{aligned}
\sigma(Y_t(i)) &= \begin{cases} 0 & \text{if no intervention for } Y_t(i) \text{ occurs} \\ 1 & \text{if intervention for } Y_t(i) \text{ occurs} \end{cases} \\
\sigma(\boldsymbol{\theta}_t(i)) &= \begin{cases} 0 & \text{if no intervention for } \boldsymbol{\theta}_t(i) \text{ occurs} \\ 1 & \text{if intervention for } \boldsymbol{\theta}_t(i) \text{ occurs} \end{cases}
\end{aligned}$$

Note that, although similar in concept, the intervention variables used here are different to those used in Dawid (2002). In that paper, atomic interventions were used with a finite set S of possible interventions, so that an intervention variable could take values $\emptyset$, indicating no intervention, or $s \in S$, indicating intervention strategy s. Here the situation is different. With random intervention the distribution for $Y_t(i)$ or $\boldsymbol{\theta}_t(i)$ can be manipulated arbitrarily at intervention. Therefore, in this paper the intervention variables, $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$, are simply indicators as to whether intervention takes place or not for $Y_t(i)$ and $\boldsymbol{\theta}_t(i)$, respectively. The following conditional distributions can then be defined, for some $\boldsymbol{h}_t(i)$, $H_t(i)$, $G_t(i)^*$ and $W_t(i)^*$.

$$\begin{aligned}
Y_t(i)|(\boldsymbol{\theta}_t(i), pa(Y_t(i)), \sigma(Y_t(i)) = 0)) &\sim (\boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i), V_t(i)), \\
Y_t(i)|(\boldsymbol{\theta}_t(i), pa(Y_t(i)), \sigma(Y_t(i)) = 1)) &\sim (\boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + h_t(i), V_t(i) + H_t(i)), \\
\boldsymbol{\theta}_t(i)|(\boldsymbol{\theta}_{t-1}(i), \sigma(\boldsymbol{\theta}_t(i)) = 0)) &\sim (G_t(i)\boldsymbol{\theta}_{t-1}(i), W_t(i)), \\
\boldsymbol{\theta}_t(i)|(\boldsymbol{\theta}_{t-1}(i), \sigma(\boldsymbol{\theta}_t(i)) = 1)) &\sim (G_t(i)^*\boldsymbol{\theta}_{t-1}(i), W_t(i)^*). \qquad (3.1)
\end{aligned}$$

The intervention variables, $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$, are not random. Instead, they are decision variables whose values are controlled by the forecaster using the forecasting model. Dawid (2002) introduced the idea of using influence diagrams for investigating causal relationships by adding intervention decision nodes into the influence diagrams so that the effects of intervention can be easily seen. Here, the same idea is used. The intervention indicator variables, $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$, will be added as decision nodes to a DBN of the MDM to produce an influence diagram of the MDM. Thus, intervention is explicitly introduced into an influence diagram of the model. The MDM uses the conditional independence structure related to causality presented in the ID. As such, the descendants of the decision variable $\sigma(Y_t(i))$ will be affected by intervention for $Y_t(i)$, and the descendants of the decision variable $\sigma(\boldsymbol{\theta}_t(i))$ will be affected by interventions for $\boldsymbol{\theta}_t(i)$. An ID therefore gives a clear picture of the effects of interventions for $Y_t(i)$ and $\boldsymbol{\theta}_t(i)$.

3.1 An influence diagram for the Multiregression Dynamic Model

In order to draw an influence diagram which represents the general structure of all MDMs, define the following notation. Let

$$\begin{aligned}
\boldsymbol{X}_t(i)^\top &= (Y_t(1), \ldots, Y_t(i-1)) \\
\boldsymbol{Z}_t(i)^\top &= (Y_t(i+1), \ldots, Y_t(n)) \\
\boldsymbol{\alpha}_t(i)^\top &= (\boldsymbol{\theta}_t(1)^\top, \ldots, \boldsymbol{\theta}_t(i-1)^\top) \\
\boldsymbol{\beta}_t(i)^\top &= (\boldsymbol{\theta}_t(i+1)^\top, \ldots, \boldsymbol{\theta}_t(n)^\top).
\end{aligned}$$

Thus, for $i = 2, \ldots, n-1$, the time series can be written as $\boldsymbol{Y}_t^\top = (\boldsymbol{X}_t(i)^\top, Y_t(i), \boldsymbol{Z}_t(i)^\top)$, with parameter vector $\boldsymbol{\theta}_t^\top = (\boldsymbol{\alpha}_t(i)^\top, \boldsymbol{\theta}_t(i)^\top, \boldsymbol{\beta}_t(i)^\top)$. Then since $pa(Y_t(i)) \subseteq \boldsymbol{X}_t(i)$, the vector $\boldsymbol{F}_t(i)$ is a known function of $\boldsymbol{X}_t(i)$, and $Y_t(i)|(\boldsymbol{X}_t(i), \boldsymbol{\theta}_t(i))$ has some distribution with mean $\boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i)$ and variance $V_t(i)$.

Suppose that $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-1}$ have been observed. An influence diagram representing the MDM before $\boldsymbol{Y}_t$ is observed, is given in Figure 3. Also included in the ID are the intervention decision variables, $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$.

Notice that there are arcs leading from D_{t-1} and $\boldsymbol{\theta}_{t-1}(i)$ to both $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$. This is due to the fact that the decision to intervene at time t is often a reflection of past behaviour of the series and/or its model. For example, when forecasting traffic flows in a road network, a lower than expected traffic flow at time $t-1$, due to an accident, for example, might be expected to be followed at time t by a higher than expected traffic flow once the accident is cleared as the vehicles who have been queuing flow through. Therefore, observing that the flow is low at time $t-1$ is likely to influence any decision to intervene at time t. Alternatively, a model monitor may indicate poor model performance at time $t-1$, and this information may influence the decision to intervene at time t. It is also possible that the decision to intervene is entirely dictated by information external to the model. An example of such a situation in the traffic flow context would be when road works are planned.

The other arcs in the ID of Figure 3 are direct consequences of the MDM as follows.

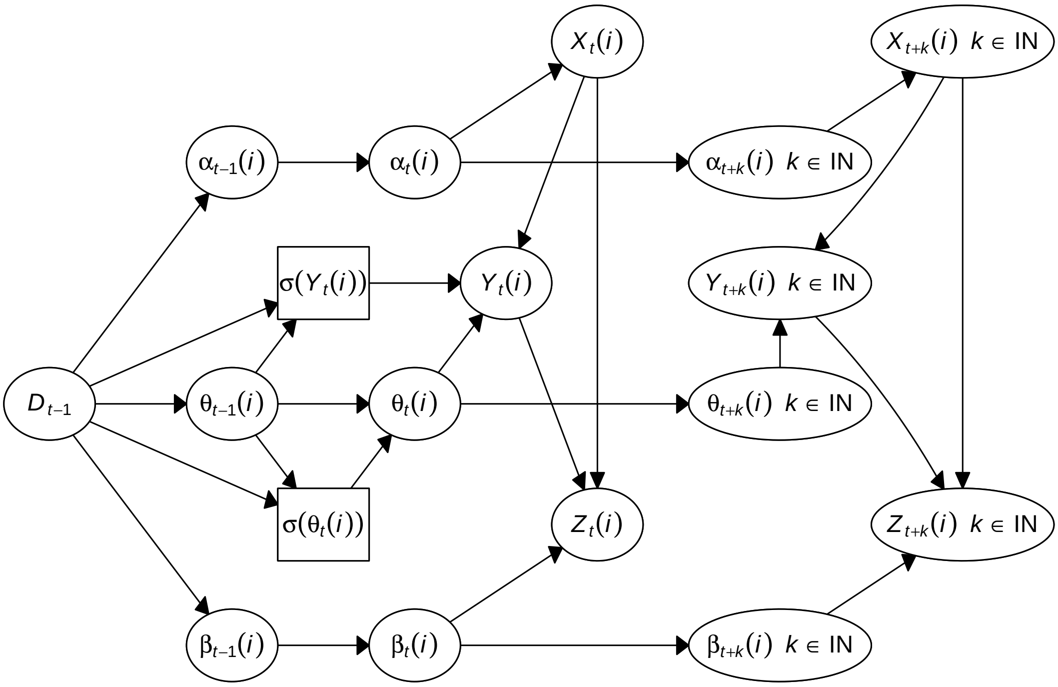

Figure 3: Influence diagram for the multiregression dynamic model before $\boldsymbol{Y}_t$ is observed, together with the intervention decision variables, $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$.

- The parameter vectors $\boldsymbol{\alpha}_{t-1}(i)$, $\boldsymbol{\theta}_{t-1}(i)$ and $\boldsymbol{\beta}_{t-1}(i)$ each have the single parent D_{t-1}, since the posterior distributions $\boldsymbol{\alpha}_{t-1}(i)|D_{t-1}$, $\boldsymbol{\theta}_{t-1}(i)|D_{t-1}$ and $\boldsymbol{\beta}_{t-1}(i)|D_{t-1}$ represent all knowledge of $\boldsymbol{\alpha}_{t-1}(i)$, $\boldsymbol{\theta}_{t-1}(i)$ and $\boldsymbol{\beta}_{t-1}(i)$, respectively, after the first $t-1$ observations are made.

- The block diagonal form of C_{t-1} ensures that $\boldsymbol{\theta}_{t-1}(1), \ldots, \boldsymbol{\theta}_{t-1}(n)$ are all mutually independent given D_{t-1} and so there are no arcs between $\boldsymbol{\alpha}_{t-1}(i)$, $\boldsymbol{\theta}_{t-1}(i)$ and $\boldsymbol{\beta}_{t-1}(i)$.

- The block diagonal forms of G_t and W_t in the system equation specify separate distributions for each $\boldsymbol{\theta}_t(i)|\boldsymbol{\theta}_{t-1}(i)$, $i = 1, \ldots, n$. Thus $\boldsymbol{\alpha}_t(i)$, $\boldsymbol{\theta}_t(i)$ and $\boldsymbol{\beta}_t(i)$ have the single parents $\boldsymbol{\alpha}_{t-1}(i)$, $\boldsymbol{\theta}_{t-1}(i)$ and $\boldsymbol{\beta}_{t-1}(i)$, respectively. As a consequence of distributions (3.1), $\boldsymbol{\theta}_t(i)$ additionally has the parent $\sigma(\boldsymbol{\theta}_t(i))$.

- The block diagonal structures of $G_{t+1}, \ldots, G_{t+k}$ and $W_{t+1}, \ldots, W_{t+k}$ ensure that repeated use of the system equation specifies separate distributions for each $\boldsymbol{\theta}_{t+k}(i)|\boldsymbol{\theta}_t(i)$, $i = 1, \ldots, n$, for $k \in \mathbb{N}$. Therefore $\{\boldsymbol{\alpha}_{t+k}(i), k \in \mathbb{N}\}$, $\{\boldsymbol{\theta}_{t+k}(i), k \in \mathbb{N}\}$ and $\{\boldsymbol{\beta}_{t+k}(i), k \in \mathbb{N}\}$ have the single parents $\boldsymbol{\alpha}_t(i)$, $\boldsymbol{\theta}_t(i)$ and $\boldsymbol{\beta}_t(i)$, respectively.

- The MDM has n separate observation equations defining the distributions for $Y_t(i)|(\boldsymbol{\theta}_t(i), pa(Y_t(i)))$, $i = 1, \ldots, n$. Thus:

  - $\boldsymbol{X}_t(i)$ has the parent $\boldsymbol{\alpha}_t(i)$,
  - $Y_t(i)$ has the parents $\boldsymbol{\theta}_t(i)$ and $\boldsymbol{X}_t(i)$,
  - $\boldsymbol{Z}_t(i)$ has the parents $\boldsymbol{\beta}_t(i)$, $\boldsymbol{X}_t(i)$ and $Y_t(i)$.

  In addition, as a consequence of distributions (3.1), $Y_t(i)$ also has the parent $\sigma(Y_t(i))$.

- The n observation equations at time $t + k$ define the distributions for $Y_{t+k}(i)|(\boldsymbol{\theta}_{t+k}(i), pa(Y_{t+k}(i)))$, $i = 1, \ldots, n$. Thus:

  - $\{\boldsymbol{X}_{t+k}(i), k \in \mathbb{N}\}$ has the parent $\{\boldsymbol{\alpha}_{t+k}(i), k \in \mathbb{N}\}$,
  - $\{Y_{t+k}(i), k \in \mathbb{N}\}$ has the parents $\{\boldsymbol{\theta}_{t+k}(i), k \in \mathbb{N}\}$ and $\{\boldsymbol{X}_{t+k}(i), k \in \mathbb{N}\}$,
  - $\{\boldsymbol{Z}_{t+k}(i), k \in \mathbb{N}\}$ has the parents $\{\boldsymbol{\beta}_{t+k}(i), k \in \mathbb{N}\}$, $\{\boldsymbol{X}_{t+k}(i), k \in \mathbb{N}\}$ and $\{Y_{t+k}(i), k \in \mathbb{N}\}$.

3.2 Effects of intervention in the MDM before observing $\boldsymbol{Y}_t$

The effects of intervention for $Y_t(i)$ and $\boldsymbol{\theta}_t(i)$ in the MDM can be easily seen by looking at the descendants of $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$, respectively, in the ID of Figure 3.

- $\sigma(Y_t(i))$ has descendants $Y_t(i)$ and $\boldsymbol{Z}_t(i)$ only. Therefore intervention for $Y_t(i)$ affects $Y_t(i)$'s forecast and also affects $\boldsymbol{Z}_t(i)$'s forecast. However, intervention for $Y_t(i)$ does not affect the forecasts for $\boldsymbol{X}_t(i)$, nor the priors for any parameters at time $t + k$, nor the k-step ahead forecasts for $\boldsymbol{Y}_{t+k}$.

- $\sigma(\boldsymbol{\theta}_t(i))$ has descendants $\boldsymbol{\theta}_t(i)$, $Y_t(i)$, $\boldsymbol{Z}_t(i)$, $\{\boldsymbol{\theta}_{t+k}(i), k \in \mathbb{N}\}$, $\{Y_{t+k}(i), k \in \mathbb{N}\}$ and $\{\boldsymbol{Z}_{t+k}(i), k \in \mathbb{N}\}$. Therefore intervention for $\boldsymbol{\theta}_t(i)$ affects:

  - the prior for $\boldsymbol{\theta}_t(i)$,
  - the one-step ahead forecasts for $Y_t(i)$ and $\boldsymbol{Z}_t(i)$,
  - the priors for $\{\boldsymbol{\theta}_{t+k}(i), k \in \mathbb{N}\}$,
  - the k-step ahead forecasts for $\{Y_{t+k}(i), k \in \mathbb{N}\}$ and $\{\boldsymbol{Z}_{t+k}(i), k \in \mathbb{N}\}$.

  It is interesting to note that intervening for $\boldsymbol{\theta}_t(i)$ does *not* affect the one-step, or k-step, ahead forecasts for $\boldsymbol{X}_t(i)$. Intervening for $\boldsymbol{\theta}_t(i)$ also does not affect priors for any parameters for components of $\boldsymbol{Y}_t$ other than $Y_t(i)$, either at time t (that is, $\boldsymbol{\alpha}_t(i)$, $\boldsymbol{\beta}_t(i)$) or at future time periods (that is, $\{\boldsymbol{\alpha}_{t+k}(i), k \in \mathbb{N}\}$, $\{\boldsymbol{\beta}_{t+k}(i), k \in \mathbb{N}\}$).

The ID given in Figure 3 is a generic representation of the structure of all MDMs. For a specific MDM, not every variable in $\boldsymbol{Z}_t(i)$ is necessarily a descendant of $Y_t(i)$. As a result, intervention for $Y_t(i)$ will not necessarily affect the forecasts for *all* variables in $\boldsymbol{Z}_t(i)$, but only the forecasts of those $Y_t(j) \in \boldsymbol{Z}_t(i)$ which are descendants of $Y_t(i)$. Similarly, intervention for $\boldsymbol{\theta}_t(i)$ will not affect the forecasts of *all* variables in $\boldsymbol{Z}_t(i)$ or $\{\boldsymbol{Z}_{t+k}(i), k \in \mathbb{N}\}$, but only those $Y_t(j)$ and $Y_{t+k}(j)$ for which $Y_t(j) \in \boldsymbol{Z}_t(i)$ are descendants of $Y_t(i)$. The following example will illustrate this.

Example 1 Consider once again the time series represented at each time t by the DAG in Figure 1. Suppose it is decided that intervention for $Y_t(2)$ is required. Then $\boldsymbol{X}_t(2) = Y_t(1)$ and $\boldsymbol{Z}_t(2)^\top = (Y_t(3), Y_t(4), Y_t(5))$.

Intervention on $Y_t(2)$ will affect the one-step forecast of $Y_t(2)$ and the one-step forecast of $\boldsymbol{Z}_t(2)$. However, the intervention will not affect all three components of $\boldsymbol{Z}_t(2)$, but just the forecasts of $Y_t(3)$ and $Y_t(5)$, since these two are the only descendants of $Y_t(2)$. The intervention will not affect forecasts for $\boldsymbol{X}_t(2) = Y_t(1)$, any future $\boldsymbol{Y}_{t+k}$, nor the priors for any current or future parameter vectors.

If instead it was decided to intervene for $\boldsymbol{\theta}_t(2)$, then this will affect the priors for $\boldsymbol{\theta}_t(2)$, the one-step forecasts for $Y_t(2)$ and $\boldsymbol{Z}_t(2)$, the prior for $\boldsymbol{\theta}_{t+k}(2)$ and the

k-step forecasts for $Y_{t+k}(2)$ and $\boldsymbol{Z}_{t+k}(2)$. However, as $Y_t(4)$ is not a descendant of $Y_t(2)$, intervention for $\boldsymbol{\theta}_t(2)$ will only affect the forecasts of the components $Y_t(3)$ and $Y_t(5)$ of $\boldsymbol{Z}_t(2)$, and $Y_{t+k}(3)$ and $Y_{t+k}(5)$ of $\boldsymbol{Z}_{t+k}(2)$. Note that the intervention will not affect the one-step forecasts for $\boldsymbol{X}_t(2) = Y_t(1)$, the k-step forecasts for $\boldsymbol{X}_{t+k}(2) = Y_{t+k}(1)$, nor the priors of any present or future parameter vectors other than $\boldsymbol{\theta}_t(2)$ and $\boldsymbol{\theta}_{t+k}(2)$. ■

3.3 Effects of intervention in the MDM after observing $\boldsymbol{Y}_t$

After observing $\boldsymbol{Y}_t$, the ID in Figure 3 is no longer appropriate to represent the MDM. Each of the arcs $(\boldsymbol{\alpha}_t(i), \boldsymbol{X}_t(i))$, $(\boldsymbol{\theta}_t(i), Y_t(i))$ and $(\boldsymbol{\beta}_t(i), \boldsymbol{Z}_t(i))$ needs to be reversed to reflect the fact that the posterior distributions $\boldsymbol{\alpha}_t(i)|\boldsymbol{Y}_t$, $\boldsymbol{\theta}_t(i)|\boldsymbol{Y}_t$ and $\boldsymbol{\beta}_t(i)|\boldsymbol{Y}_t$ are now of interest (rather than the distributions $Y_t(i)|\boldsymbol{\theta}_t(i)$ specified by the observation equations used for forecasting each $Y_t(i)$). Following Howard and Matheson's Arc Reversal Theorem (Howard and Matheson, 1984), extra arcs need to be introduced into the ID. Explicitly, reversing the arc between any two nodes A and B means that A must inherit B's parents and B must inherit A's parents. The new influence diagram, after observing $\boldsymbol{Y}_t$, is given in Figure 4.

It is important to note that the interventions for $Y_t(i)$ and/or $\boldsymbol{\theta}_t(i)$ still precede the observation $\boldsymbol{Y}_t$. However, as the ID representing the MDM changes after observing $\boldsymbol{Y}_t$, so the effects of the interventions will change after $\boldsymbol{Y}_t$ is observed. The effects of intervention after $\boldsymbol{Y}_t$ is observed are easily seen from the ID in Figure 4 by again looking at the descendants of $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$, respectively. This time both $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$ have the *same* descendants: $Y_t(i)$, $\boldsymbol{\theta}_t(i)$, $\boldsymbol{Z}_t(i)$, $\boldsymbol{\beta}_t(i)$, $\{\boldsymbol{\theta}_{t+k}(i), k \in \mathbb{N}\}$, $\{\boldsymbol{\beta}_{t+k}(i), k \in \mathbb{N}\}$, $\{Y_{t+k}(i), k \in \mathbb{N}\}$ and $\{\boldsymbol{Z}_{t+k}(i), k \in \mathbb{N}\}$. Thus after observing $\boldsymbol{Y}_t$, intervention for $Y_t(i)$ and $\boldsymbol{\theta}_t(i)$ both affect

- the posterior for $\boldsymbol{\theta}_t(i)$ and also the posterior for $\boldsymbol{\beta}_t(i)$,
- the priors for future $\boldsymbol{\theta}_{t+k}(i)$ and also the priors for future $\boldsymbol{\beta}_{t+k}(i)$,
- the k-step forecasts for $Y_{t+k}(i)$ and the k-step forecasts for $\boldsymbol{Z}_{t+k}(i)$.

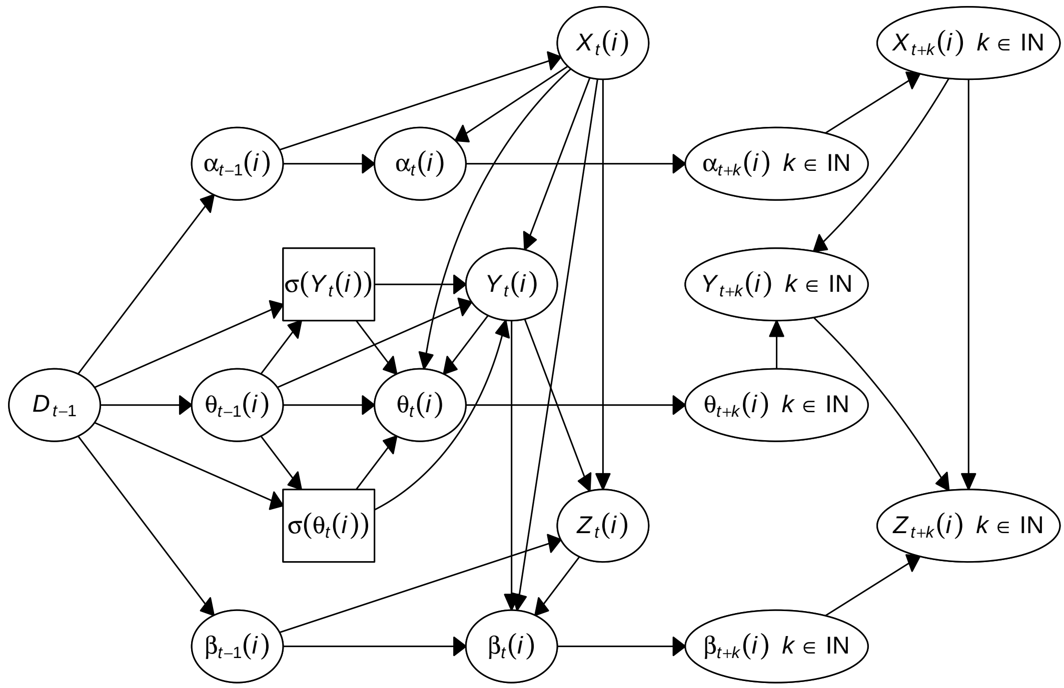

Figure 4: Influence diagram for the multiregression dynamic model, together with intervention indicator decision variables $\sigma(Y_t(i)$ and $\sigma(\boldsymbol{\theta}_t(i))$, *after* $\boldsymbol{Y}_t$ is observed.

Neither intervention for $Y_t(i)$ nor $\boldsymbol{\theta}_t(i)$ affect the posterior for $\boldsymbol{X}_t(i)$'s parameter vector $\boldsymbol{\alpha}_t(i)$, the prior for future parameters $\boldsymbol{\alpha}_{t+k}$, nor the k-step forecasts for $\boldsymbol{X}_{t+k}(i)$.

Notice that before $\boldsymbol{Y}_t$ is observed, intervening for $Y_t(i)$ affects a different set of variables than intervening for $\boldsymbol{\theta}_t(i)$ does. On the other hand, after $\boldsymbol{Y}_t$ is observed, intervening for $Y_t(i)$ affects exactly the same set of variables (although in different ways) as intervening for $\boldsymbol{\theta}_t(i)$. It is interesting to note that neither intervention affects the forecasts of $\{\boldsymbol{X}_{t+k}(i), k \in \mathbb{N}\}$, nor the distributions of its parameters.

Example 2 Consider once again the time series represented at time t by the DAG in Figure 1. Example 1 looked at the effects of intervention for $Y_t(2)$ and $\boldsymbol{\theta}_t(2)$ before $\boldsymbol{Y}_t$ is observed. Now suppose that $\boldsymbol{Y}_t$ is observed. Then intervention for either $Y_t(2)$ or $\boldsymbol{\theta}_t(2)$ will affect the same distributions (though not in the same way).

As in Example 1, $\boldsymbol{X}_t(2) = Y_t(1)$ and $\boldsymbol{Z}_t(2)^\top = (Y_t(3), Y_t(4), Y_t(5))$. Also $\boldsymbol{\alpha}_t(2) =$

$\boldsymbol{\theta}_t(1)$ and $\boldsymbol{\beta}_t(2)^\top = (\boldsymbol{\theta}_t(3)^\top, \boldsymbol{\theta}_t(4)^\top, \boldsymbol{\theta}_t(5)^\top)$. Intervention for $Y_t(2)$ or $\boldsymbol{\theta}_t(2)$ will affect the posterior for $\boldsymbol{\theta}_t(2)$, the prior for $\boldsymbol{\theta}_{t+k}(2)$, and the k-step forecast for $Y_{t+k}(2)$. The interventions will also affect the posterior for $\boldsymbol{\beta}_t(2)$, the prior for $\boldsymbol{\beta}_{t+k}(2)$ and the k-step forecast for $\boldsymbol{Z}_{t+k}(2)$. However, as $Y_t(4)$ is not a descendant of $Y_t(2)$, intervention for $Y_t(2)$ or $\boldsymbol{\theta}_t(2)$ will not affect all three components of $\boldsymbol{\beta}_t(2)$, $\boldsymbol{\beta}_{t+k}(2)$ and $\boldsymbol{Z}_{t+k}(2)$. Instead it will only affect the posteriors for $\boldsymbol{\theta}_t(3)$ and $\boldsymbol{\theta}_t(5)$, the priors for $\boldsymbol{\theta}_{t+k}(3)$ and $\boldsymbol{\theta}_{t+k}(5)$, and the k-step forecasts for $Y_{t+k}(3)$ and $Y_{t+k}(5)$.

Neither intervention will affect the posterior for $\boldsymbol{\alpha}_t(2) = \boldsymbol{\theta}_t(1)$, the prior for $\boldsymbol{\alpha}_{t+k}(2) = \boldsymbol{\theta}_{t+k}(1)$, nor the k-step forecast for $\boldsymbol{X}_{t+k}(2) = Y_{t+k}(1)$. ■

4 Intervention for road traffic flow forecasting

In this section, intervention in the MDM will be illustrated using a multivariate time series of road traffic flows. Two different interventions will be considered — one for an observed series $Y_t(i)$ and the other one for two state vectors.

In Queen *et al.* (2007a) the LMDM was used to forecast traffic flows at a junction of three major roads in the UK. Hourly vehicle counts were recorded at a number of data collection sites across the traffic network. Each data collection site is labelled by a number and the vehicle count for hour t at site s is denoted by $Y_t(s)$. The possible routes through the network were used to elicit an ID suitable for use with an MDM. The ID elicited for part of the network representing vehicle counts at sites labelled 167, 168, 170A, 170B, 169, 161 and 171 is shown in Figure 5. (For full details regarding the ID and how it was elicited, see Queen *et al.* (2007a).) In the ID of Figure 5, $Y_t(168)$, $Y_t(170\text{A})$ and $Y_t(171)$ are logical functions of their parents (explicitly $Y_t(168) = Y_t(167) - (Y_t(170\text{A}) + Y_t(170\text{B}))$, $Y_t(170\text{A}) = (Y_t(170\text{A}) + Y_t(170\text{B})) - Y_t(170\text{B})$ and $Y_t(171) = (Y_t(161) + Y_t(171)) - Y_t(161)$). Following the terminology of WinBUGS software (http://www.mrc-bsu.cam.ac.uk/bugs/), these are called logical variables and denoted by double ovals. The variables $Z_t(1)$ and $Z_t(2)$ are also logical variables. These are simply two orthogonal variables which are functions of $Y_t(170\text{B})$ and $Y_t(169)$. $Z_t(1)$ and $Z_t(2)$ were created to overcome problems of collinearity between $Y_t(170\text{B})$ and $Y_t(169)$ and can essentially be ignored in this paper.

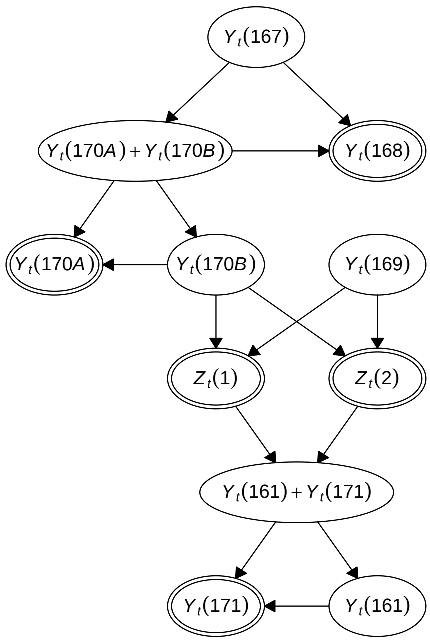

Figure 5: Part of the influence diagram for each time t (taken from Queen *et al.* (2007a)) representing vehicle counts at a junction of three major roads in the UK.

The daily pattern of traffic flows varies depending on the day of the week. However, the pattern is similar for Tuesday, Wednesday and Thursday each week. For clarity when illustrating intervention in the MDM, only traffic flows for Tuesday-Thursday each week are considered here.

4.1 Intervention for a $Y_t(i)$

Poor forecast performance, as identified by large forecast errors, can be an indication that intervention might be useful. Such an instance of poor forecast performance occurs in series $Y_t(167)$, where an unusually large negative forecast error occurs at time 560 (07.00-08.00 on Thursday week 8), followed by an unusually large positive forecast error. This pattern of forecast errors is consistent with a slowdown in traffic flow, for example due to a temporary block in the road following a crash, followed by an increase in traffic flow as the problem is resolved and delayed cars move through the network. Such patterns are not uncommon in traffic networks. A plot of the one-step ahead forecast errors with ± 1.96 forecast standard deviation error bars for $Y_t(167)$ between times 500 and 580 is shown in Figure 6(a). The forecast errors for times 560 and 561 are circled on the plot. The same pattern of forecast errors is also evident in $Y_t(167)$'s children (see for example Figure 6(c)), and in other descendants (see for example the plot of the one-step forecast errors for $Y_t(167)$'s 'grandchild' $Y_t(170\text{B})$ shown in Figure 6(e)). This is essentially due to the fact that traffic from site 167 flows to sites 168, 170A or 170B, so any changes in traffic flow at site 167 will have a knock-on effect to the traffic flows at these sites, and, in turn, to the flows at sites further downstream. It is precisely these kinds of relationships between flows at different sites in the network which the ID was designed to represent. Figure 6(g) shows a plot of the one-step forecast errors for $Y_t(169)$ over the same time period. It is interesting to note that $Y_t(169)$ is a non-descendant of $Y_t(167)$ and does not show the same large forecast errors at times 560 and 561.

In order to improve forecast performance of $Y_t(167)$ and its descendants, intervention was used for $Y_t(167)$ at time 561, following the unexpectedly large negative forecast error at time 560. Without intervention, the observation equation for $t = 561$ gives the distribution:

$$Y_t(167)|(\boldsymbol{\theta}_t(167), \sigma(Y_t(167)) = 0) \sim N(\boldsymbol{F}_t(167)^\top \boldsymbol{\theta}_t(167), V_t(167)).$$

Intervention was done as follows. The observation $y_{560}(167)$ was unexpected and so was treated simply as an outlier. During the following time period ($t = 561$), as the road blockage clears and vehicles start moving, the delayed vehicles (from the previous hour) are expected to pass site 167, in addition to the vehicles that arrive during hour $t = 561$. The expected number of cars delayed from hour 560 is $e_{560}(167) = f_{560}(167) - y_{560}(167)$, where $f_{560}(167)$ is the one-step forecast (at time 559) for $Y_{560}(167)$. The observation error variance is also increased at time 561 to allow for increased uncertainty. Thus intervention adjusts the observation equation

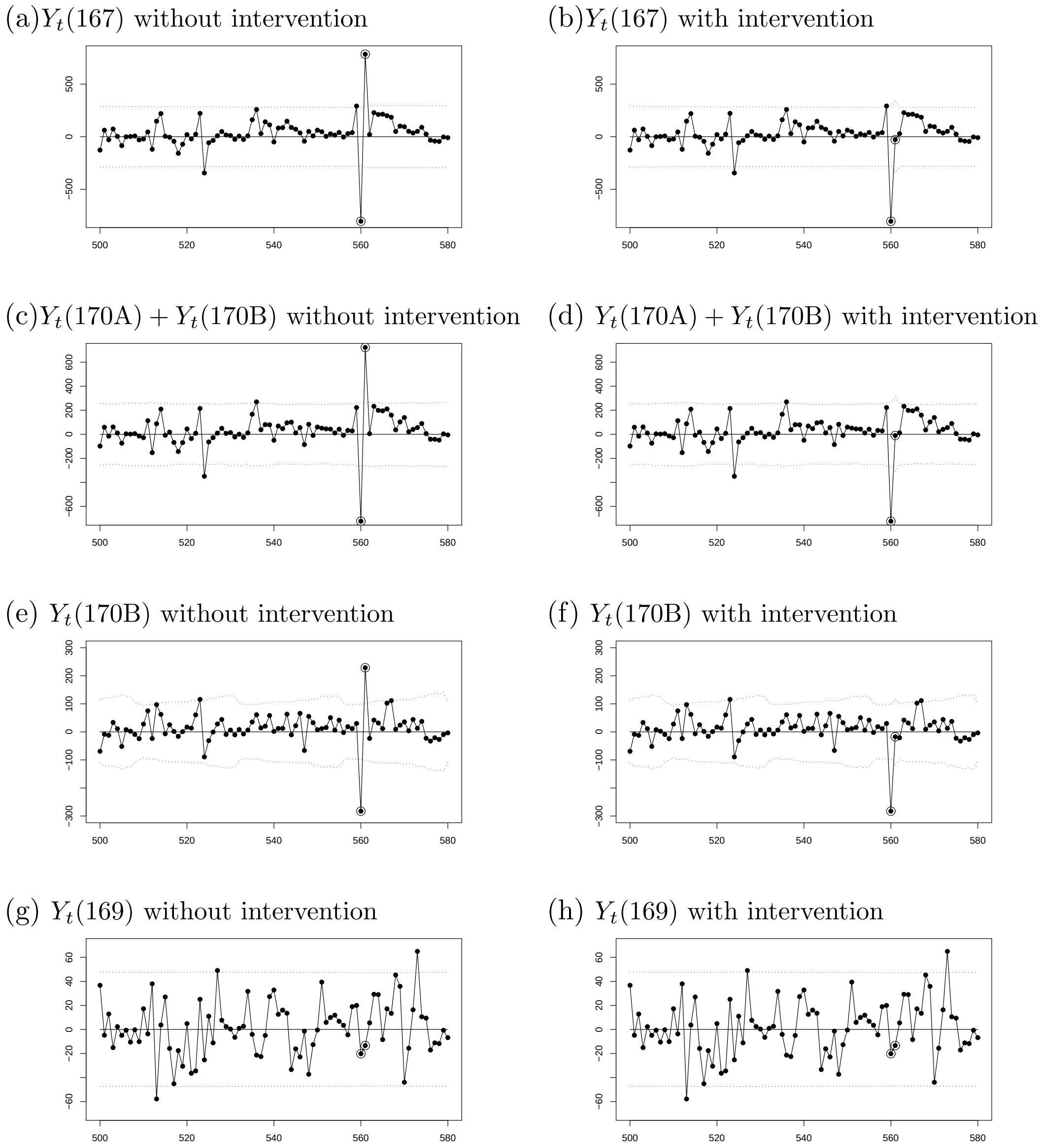

Figure 6: Plots of the one-step forecast errors (solid line) and ± 1.96 forecast standard deviations (dotted lines) obtained using the Linear Multiregression Dynamic Model between times 500 and 580 for $Y_t(167)$, one of its children $Y_t(170\text{A}) + Y_t(170\text{B})$, one of its 'grandchildren' $Y_t(170\text{B})$ and a non-descendant $Y_t(169)$. Plots on the left are the forecast errors obtained without intervention and those on the right are the forecast errors following intervention for $Y_{561}(167)$. The observations at times 560 and 561 are circled for each series in each plot.

at hour 561 so that for $t = 561$,

$$Y_t(167) = \boldsymbol{F}_t(167)^\top \boldsymbol{\theta}_t(167) + e_{t-1}(167) + v_t(167), \qquad v_t(167) \sim N(0, V_t(167) + 10000),$$

giving the distribution:

$$Y_t(167)|(\boldsymbol{\theta}_t(167), \sigma(Y_t(167)) = 1) \sim N(\boldsymbol{F}_t(167)^\top \boldsymbol{\theta}_t(167) + e_{t-1}(167), V_t(167) + 10000).$$

Section 3.2 looked at the effects of an intervention for $Y_t(i)$ before $Y_t(i)$ is observed. From the ID of the MDM it was shown that such an intervention will affect the one-step forecast for $Y_t(i)$ and also its descendants' forecasts, but will not affect the one-step forecasts for any non-descendants. The effects of intervening for $Y_{561}(167)$ on the one-step ahead forecasts can be clearly seen in the right hand plots of Figure 6 — the intervention not only improves the one-step forecast error for $Y_{561}(167)$ (plot (b)), but it also improves the one-step forecast error for its children (see plot (d)), its 'grandchildren' (see plot (f)) and indeed its 'great grandchildren' (not shown). On the other hand, it is clearly seen that the intervention for $Y_{561}(167)$ has no affect on the one-step forecast error of $Y_{561}(169)$ (plot (h)), a non-descendant of $Y_t(167)$.

In Section 3.2 it was shown that an intervention for $Y_t(i)$ will not affect any k-step ahead forecasts before $Y_t(i)$ is observed. In contrast, in Section 3.3 it was shown that, after $Y_t(i)$ is observed, the intervention *does* affect the k-step forecasts for $Y_t(i)$ and its descendants, but not the k-step forecasts of any non-descendants. To demonstrate the effect the intervention for $Y_{561}(167)$ has on the k-step ahead forecasts, Tables 1 and 2 show the k-step forecast moments for some of the series at time $t = 585$ (24 hours later).

Table 1 shows 25-step forecast moments made after $\boldsymbol{Y}_{560}$ has been observed and after any intervention for $Y_{561}(167)$ has been done, but before observing $\boldsymbol{Y}_{561}$. The first two columns show the 25-step forecast moments when there is no intervention and the last two columns show the results when there is intervention. As expected from Section 3.2, the k-step forecast distributions are exactly the same for all the series regardless of whether intervention was used or not.

Series	No intervention $f_{585}(25)$	$Q_{585}(25)$	Intervention $f_{585}(25)$	$Q_{585}(25)$
$Y_t(167)$	4,652.5	23,063.9	4,652.5	23,063.9
$Y_t(168)$	325.8	7,604.6	325.8	7,604.6
$Y_t(170\text{A}) + Y_t(170\text{B})$	4,212.0	3,688.0	4,212.0	3,688.0
$Y_t(170\text{B})$	1,430.8	5,706.8	1,430.8	5,706.8
$Y_t(161)$	1,617.6	91,114.3	1,617.6	91,114.3
$Y_t(169)$	513.9	654.0	513.9	654.0

Table 1: The 25-step forecast mean, $f_{585}(25)$, and the 25-step forecast variance, $Q_{585}(25)$, with and without intervention for $Y_{561}(167)$. These forecasts were made after $\boldsymbol{Y}_{560}$ has been observed and after any intervention has been done, but before observing $\boldsymbol{Y}_{561}$.

Series	No intervention $f_{585}(24)$	$Q_{585}(24)$	Intervention $f_{585}(24)$	$Q_{585}(24)$
$Y_t(167)$	4,356.5	21,429.9	5,456.6	38,018.1
$Y_t(168)$	398.6	9,127.7	469.8	12,356.3
$Y_t(170\text{A}) + Y_t(170\text{B})$	4,111.5	3,506.8	4,845.6	4,767.5
$Y_t(170\text{B})$	1,372.2	5,302.1	1,617.3	7,211.1
$Y_t(161)$	1,924.7	113,521.8	2,059.5	128,917.6
$Y_t(169)$	506.5	575.7	506.5	575.7

Table 2: The 24-step forecast mean, $f_{585}(24)$, and the 24-step forecast variance, $Q_{585}(24)$, with and without intervention for $Y_{561}(167)$. These forecasts were made after any intervention has been done, and after observing $\boldsymbol{Y}_{561}$.

Table 2 shows 24-step forecast moments made after any intervention for $Y_{561}(167)$ has been done, and after observing $\boldsymbol{Y}_{561}$. The first two columns show the 24-step forecast moments when there is no intervention and the last two columns show the results after intervening for $Y_{561}(167)$. As expected from Section 3.3, this time it is clearly evident that the intervention does affect k-step forecasts as the moments are quite different for $Y_t(167)$ and all its descendants (even as far away as $Y_t(161)$). Notice, that also as expected, the k-step forecast moments for non-descendant $Y_t(169)$ are the same regardless of whether intervention took place or not.

From Figure 5, it can be seen that $Y(167)$ does not have any parents. As a consequence, the MDM models $Y_t(167)$ by any suitable univariate DLM. Thus the effects of intervention for $Y_{561}(167)$ within the DLM are also illustrated. In particular, it can be clearly seen that before $\boldsymbol{Y}_{561}$ is observed, intervention for $Y_{561}(167)$ does not affect the forecasts for $Y_{585}(167)$. However, after $\boldsymbol{Y}_{561}$ is observed, intervention does affect the forecast distributions for $Y_{585}(167)$.

4.2 Intervention for two state vectors

The second intervention considered involves the state vectors for $Y_t(161) + Y_t(171)$ and for $Y_t(161)$, which will be denoted here as $\boldsymbol{\theta}_t(161.171)$ and $\boldsymbol{\theta}_t(161)$, respectively. The series $Y_t(161)+Y_t(171)$ is the number of vehicles leaving a motorway, the M25, to join another major road, the A2, at hour t. The series $Y_t(171)$ is then the number of these vehicles who travel eastbound on the A2, and series $Y_t(161)$ is the number who travel westbound.

From time 265 until time 360 (a period of four days) there was a reduction in the number of vehicles leaving the M25 to join the A2 (i.e. $Y_t(161) + Y_t(171)$) and a reduction in the number of vehicles who travelled eastbound (i.e. $Y_t(171)$). The number of vehicles who travelled westbound (i.e. $Y_t(161)$) remained, however, at the same level. This sort of pattern in traffic flow is consistent with the expected consequences of roadworks eastbound on the A2, where drivers are forewarned of this on the M25 so that fewer vehicles leave the M25 to join the A2 eastbound. This can be seen on the plot of the two series $Y_t(161)$ (in dark grey) and $Y_t(171)$ (on top in light grey) given in Figure 7.

Intervention is required in order to accommodate these reductions in flows. As the reductions in flows are persistent over a period of time (rather than for just one or two time points), intervention for the state vectors associated with $Y_{265}(161) + Y_{265}(171)$ and $Y_{265}(171)$ would be appropriate. However, it can be seen from Figure 5 that $Y_t(171)$ is a logical variable, which is calculated from $(Y_t(161) + Y_t(171)) - Y_t(161)$. So intervening for the parameters associated with $Y_{265}(161) + Y_{265}(171)$ and $Y_{265}(171)$ is equivalent to intervening for the parameters associated with $Y_{265}(161) + Y_{265}(171)$ and $Y_{265}(161)$. Thus, intervention for

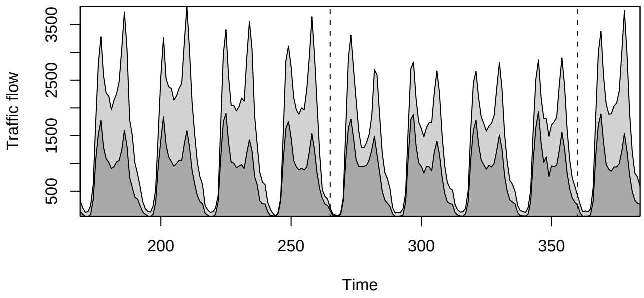

Figure 7: Plot of $Y_t(161)$ (in dark grey) and $Y_t(171)$ (on top in light grey) between times 169 and 384 (exactly 4 days before and 1 day after the period of change). Dotted vertical lines mark the period of change.

$\boldsymbol{\theta}_{265}(161.171)$ and $\boldsymbol{\theta}_{265}(161)$ is required.

The idea behind the intervention at time $t = 265$, is to scale the mean for $\boldsymbol{\theta}_{265}(161.171)$ by some value α, for $0 < \alpha < 1$, and, because $Y_t(161)$ does not change at time 265, to scale the mean for $\boldsymbol{\theta}_{265}(161)$ by $1/\alpha$. At time $t = 360$, traffic flows return to their pre-intervention levels, so a further intervention is required scaling the mean for $\boldsymbol{\theta}_{360}(161.171)$ by $1/\alpha$ and the mean for $\boldsymbol{\theta}_{360}(161)$ by α.

When using intervention for planned roadworks at time 265, ideally expert information should be used to estimate α. However, unfortunately no expert information was available for these data. It is possible that a prior could be placed on α, and α could then be estimated on-line from the data. For simplicity though, in order to illustrate the affect of the intervention *as if* good expert information were available, here the data between times 265 and 360 are used to estimate α. Using these data, the flows for $Y_t(171)$ roughly decrease by 1/3 from time 265. So, since the traffic flows for $Y_t(161)$ and $Y_t(171)$ are similar, this suggests an estimate for α of 5/6. Uncertainty concerning the intervention is incorporated into the model by increasing the variances for $\boldsymbol{\theta}_{265}(161.171)$ and $\boldsymbol{\theta}_{265}(161)$. These are regression parameters and, as such, are not expected to vary greatly over time and so, because of the scale of the system error variance, are only increased by 0.01 and 0.005, respectively. The resulting intervention distributions for time $t = 265$ are as follows.

$$\boldsymbol{\theta}_t(161.171)|(\boldsymbol{\theta}_{t-1}(161.171), \sigma(\boldsymbol{\theta}_t(161.171)) = 1)$$
$$\sim N\left(5/6 \times G_t(161.171)\boldsymbol{\theta}_{t-1}(161.171), W_t(161.171) + 0.01I\right),$$
$$\boldsymbol{\theta}_t(161)|(\boldsymbol{\theta}_{t-1}(161), \sigma(\boldsymbol{\theta}_t(161)) = 1) \sim N\left(6/5 \times G_t(161)\boldsymbol{\theta}_{t-1}(161), W_t(161) + 0.005I\right),$$

where $G_t(161.171)$, $G_t(161)$, $W_t(161.171)$ and $W_t(161)$ are the pre-intervention matrices (as in distributions (3.1)), and I is the identity matrix.

In Section 3.2, it was shown how before $Y_t(i)$ is observed, an intervention for $\boldsymbol{\theta}_t(i)$ will affect the one-step forecasts for $Y_t(i)$ and its descendants, but not affect any one-step forecasts of non-descendants. Figure 8 shows the one-step forecast errors both with intervention for $\boldsymbol{\theta}_{265}(161.171)$ and $\boldsymbol{\theta}_{265}(161)$ (solid line) and without any intervention (dotted line) for several series. From this it is clearly seen how the one-step forecasts are affected by the interventions on the state vectors for $Y_t(161) + Y_t(171)$ (plot (a)) and the descendant $Y_t(171)$ (plot (c)). The one-step forecasts for $Y_t(161)$ are also affected, but to a much lesser extent (see plot (b)). This is because $Y_t(161)$ did not exhibit any change at the intervention period and so a change in its forecasts and forecast errors were not actually required. As expected, these interventions have absolutely no effect on the one-step forecasts for non-descendant $Y_t(170\text{A}) + Y_t(170\text{B})$ (plot (d)).

In contrast to intervention for $Y_t(i)$ which does not affect any k-step forecasts before $\boldsymbol{Y}_t$ is observed, but does after $\boldsymbol{Y}_t$ is observed, intervention for $\boldsymbol{\theta}_t(i)$ affects the k-step forecasts of $Y_t(i)$ and its descendant both before *and* after $\boldsymbol{Y}_t$ is observed. This is demonstrated in Tables 3 and 4 which show the k-step forecast moments for $Y_t(161) + Y_t(171)$ and $Y_t(161)$ at time $t = 289$ (24 hours after the intervention at time 265).

Table 3 shows the 25-step forecast moments for time 289 made after Y_{264} has been observed and after any interventions for $\boldsymbol{\theta}_{265}(161.171)$ and $\boldsymbol{\theta}_{265}(161)$ have been done, but before $\boldsymbol{Y}_{265}$ is observed. Table 4 shows the 24-step forecast moments for time 289 made after any interventions for $\boldsymbol{\theta}_{265}(161.171)$ and $\boldsymbol{\theta}_{265}(161)$ have been done, and after $\boldsymbol{Y}_{265}$ is observed. For both tables, the first two columns show the

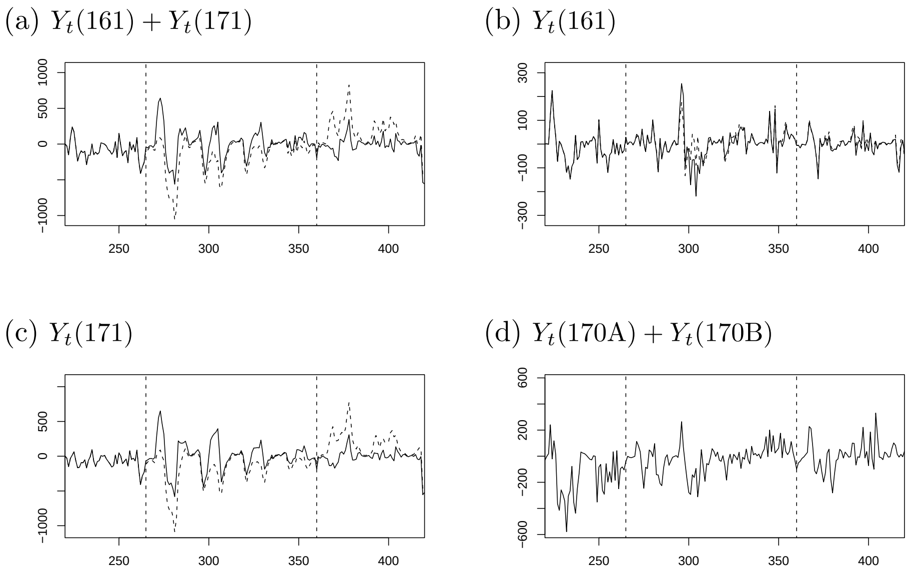

Figure 8: The one-step forecast errors both with intervention for $\boldsymbol{\theta}_{265}(161.171)$ and $\boldsymbol{\theta}_{265}(161)$ (solid line) and without any intervention (dotted line) for $Y_t(161){+}Y_t(171)$, $Y_t(161)$, their descendant $Y_t(171)$ and non-descendant $Y_t(170\text{A}) + Y_t(170\text{B})$. Dotted vertical lines mark the period of change and the intervention takes place at the first dotted line.

Series	No intervention $f_{289}(25)$	$Q_{289}(25)$	Intervention $f_{289}(25)$	$Q_{289}(25)$
$Y_t(161) + Y_t(171)$	526.4	351,407.4	433.7	355,004.0
$Y_t(161)$	317.9	7,117.6	385.8	11,008.1

Table 3: The 25-step forecast mean, $f_{289}(25)$, and the 25-step forecast variance, $Q_{289}(25)$, with and without intervention for $\boldsymbol{\theta}_{265}(161.171)$ and $\boldsymbol{\theta}_{265}(161)$. These forecasts were made after $\boldsymbol{Y}_{264}$ has been observed and after any interventions have been done, but before $\boldsymbol{Y}_{265}$ is observed.

Series	No intervention $f_{289}(24)$	$Q_{289}(24)$	Intervention $f_{289}(24)$	$Q_{289}(24)$
$Y_t(161) + Y_t(171)$	600.6	205,815.9	494.9	210,372.2
$Y_t(161)$	294.4	5,532.9	357.2	8,314.3

Table 4: The 24-step forecast mean, $f_{289}(24)$, and the 24-step forecast variance, $Q_{289}(24)$, with and without intervention for $\boldsymbol{\theta}_{265}(161.171)$ and $\boldsymbol{\theta}_{265}(161)$. These forecasts were made after any interventions have been done, and after $\boldsymbol{Y}_{265}$ is observed.

k-step forecast moments when there is no intervention and the last two columns show the results when there is intervention. From these tables, the effects of the interventions on k-step forecasts both before $\boldsymbol{Y}_{265}$ is observed and after $\boldsymbol{Y}_{265}$ is observed are clearly seen. Notice that the interventions will also affect the k-step forecasts both before and after $\boldsymbol{Y}_{265}$ is observed for the descendant $Y_t(171)$ as this is simply a logical function of its parents.

5 Identification of causal relationships between time series

The MDM is defined by the DBN representing the conditional independence structure related to causality across the time series over time. In particular, the BN for the time series at time t represents contemporaneous causal relationships between the component series. The forecast performance of an MDM is therefore informative about these assumed contemporaneous causal relationships.

As mentioned in Section 2, two BNs can represent the same conditional independence statements, but have quite different conditional independence structures related to causality, and hence quite different MDMs. Suppose that there are several possible MDMs for $\boldsymbol{Y}_t$, each of which has the same conditional independence structure, but each with a different conditional independence structure related to causality. Following the ideas of multiprocess DLMs (Harrison and Stevens, 1976), each of these competing MDMs can be modelled simultaneously using a *multiprocess MDM*. In a multiprocess MDM, each competing MDM has an associated probability of being the 'correct' model. Therefore, for each competing MDM, there is a probability that the corresponding BN represents the 'correct' contemporaneous causal relationships. As data are observed, these probabilities are updated. Thus a multiprocess MDM can provide on-line assessments of causal relationships between time series.

As is well-known from BN and DBN theory, without intervention it can be extremely difficult to identify which causal relationships are 'correct' from a set of competing models with the same conditional independence structure. As a consequence, at times when there is no intervention, there should be little difference between the various model probabilities in a multiprocess MDM. However, this is not the case when intervention is used. As it was shown in Section 3, intervention for $Y_t(i)$ (or $\boldsymbol{\theta}_t(i)$) will affect the distributions of its descendants. Therefore any MDM which continues to provide good forecasts for the descendants of $Y_t(i)$ is likely to be representing the 'correct' causal structure, and the associated probability for that model will be larger than the others.

Explicitly, suppose there are m competing BNs for $\boldsymbol{Y}_t$ with associated MDMs $M_1, \ldots, M_m$. Let $p_{t-1}(j)$ denote the prior probability that model M_j is 'correct' given D_{t-1}, for $j = 1, \ldots, m$. Then after observing $\boldsymbol{Y}_t = \boldsymbol{y}_t$, the posterior probability can be calculated via

$$p_t(j) = c_t p_{t-1}(j) l_t(j),$$

where c_t is the normalising constant and $l_t(j)$ is the likelihood for $\boldsymbol{y}_t$ under model M_j. In this case the likelihood function is the observed value of the one-step forecast distribution $f(\boldsymbol{y}_t|D_{t-1}, M_j)$, which in the MDM is the product of the observed values of individual univariate conditionals. The posterior probability that M_j is 'correct' is thus given by

$$p_t(j) = c_t p_{t-1}(j) \prod_{i=1}^{n} f(y_t(i)|pa(y_t(i)), D_{t-1}, M_j).$$

The model probabilities are therefore updated over time and a decision at time t as to which causal structure is 'correct' is based on the posterior probabilities $p_t(1), \ldots, p_t(m)$.

To illustrate using a multiprocess MDM to identify a contemporaneous causal relationship, consider once again the traffic network introduced in Section 4. Suppose for simplicity that there are only the following two competing models.

- The MDM already used in Section 4 with ID given in Figure 5. Call this model M_1.
- An MDM using an ID which is the same as in Figure 5 except that the arc between $Y_t(167)$ and $Y_t(170\text{A}) + Y_t(170\text{B})$ is reversed. Call this model M_2.

Notice that the IDs for both models represent the same conditional independence structure, but represent different conditional independences related to causality. As there are only two possible models here, the ratio of the posterior model probabilities, $p_t(1)/p_t(2)$ is of particular interest.

In order to illustrate how the multiprocess MDM can identify causal relationships, the ratio of the posterior model probabilities is calculated at two separate times — at time 540 when no intervention is required, and at time 561 following the intervention for $Y_{561}(167)$ as detailed in Section 4.1. Time $t = 540$ was chosen fairly randomly to illustrate the value of the ratio of posterior model probabilities when the model is performing well and no intervention is required. In Section 4.1, model M_1 was used and it was shown how intervention for $Y_{561}(167)$ has a dramatic effect on the one-step forecasts and resulting forecast errors of $Y_{561}(167)$ and its descendants (see Figure 6). Because $Y_t(167)$ has so many descendants in the ID for

M_1, the effects of the intervention for $Y_{561}(167)$ are seen across almost the entire network. In model M_2, however, $Y_t(167)$ only has the single descendant $Y_t(168)$ and so intervention for $Y_{561}(167)$ will have only a limited affect on the one-step forecasts over the network as a whole. In particular, intervention for $Y_{561}(167)$ will have no affect in reducing the large one-step forecast errors for $Y_{561}(170\text{A}) + Y_{561}(170\text{B})$ and its descendants. The effects of the intervention are therefore quite different for the two models and the observed likelihoods, $l_t(1)$ and $l_t(2)$, and hence the posterior probabilities, $p_t(1)$ and $p_t(2)$, will reflect this.

The initial two weeks of data were used to estimate priors for the parameters for both models. The models were then run (separately) and updated sequentially as usual up to time $t = 539$. For fairness, at time 539 both models are assigned the same prior probability so that $p_{539}(1) = p_{539}(2) = 0.5$. After observing $\boldsymbol{Y}_{540} = \boldsymbol{y}_{540}$, the ratio of posterior probabilities, $p_{540}(1)/p_{540}(2)$, is calculated to be 1.06. This is approximately equal to 1 as expected, illustrating how it is difficult to identify causality when no intervention is used. The models were then again run (separately) and updated sequentially as usual up to time $t = 560$. At time $t = 561$, intervention was performed for $Y_{561}(167)$ using the same method as presented in Section 4.1 so that the distribution for $Y_{561}(167)|(\boldsymbol{\theta}_{561}(167), M_j)$ was adjusted at intervention by adding $f_{560}(167) - y_{560}(167)$ to the mean and adding 10,000 to the variance. For fairness, at time 560 again both models are assigned the same prior probability, so that $p_{560}(1) = p_{560}(2) = 0.5$. After observing $\boldsymbol{Y}_{561} = \boldsymbol{y}_{561}$, the ratio of posterior probabilities, $p_{561}(1)/p_{561}(2)$, is this time calculated to be 1.7×10^{30} providing overwhelming support for model M_1 as opposed to model M_2. Notice how the ratio of posterior probabilities is much larger at the time of intervention ($t = 561$), than at the time of no intervention ($t = 540$), thus illustrating how much easier it is to identify causal relationships at the time of an intervention then when there is no intervention. For this particular application, the context of the problem heuristically suggests that $Y_t(167)$ is causal for $Y_t(170\text{A}) + Y_t(170\text{B})$, since traffic flows from site 167 to sites 170A and 170B. It is therefore reassuring that the multiprocess MDM so clearly confirms this when intervention is used.

It is possible that causal relationships in a multivariate series may change over time. For example, in the traffic network application, a queue of vehicles at one site can cause a slowdown in traffic at upstream sites, thus temporarily reversing the causality between sites. This is, in fact, an important modelling issue when forecasting traffic flows in practice. To accommodate changes in causal relationships, multiprocess MDMs can be used at each time point to provide an on-line assessment of the most likely causal relationships at each time t. Further investigation of this will be the focus of future research.

6 Concluding remarks

This paper has shown how intervention can be an extremely useful tool for forecasting in the MDM. The effects of intervention, for both the observed series $Y_t(i)$ and the underlying state vectors $\boldsymbol{\theta}_t(i)$, have been explored and have been illustrated for multivariate traffic flow series. The idea of using multiprocess MDMs for identifying contemporaneous causal relationships between time series has been introduced, going beyond previous research using intervention for identifying causality between lagged time series. A simple multiprocess MDM was used with the traffic flow series to identify the more likely causal structure from two competing alternatives, both at a time when the model was performing well and no intervention was required, and at the time of an intervention. The respective values of the resulting ratio of posterior model probabilities demonstrated just how effective the multiprocess model is at identifying contemporaneous causal relationships at a time of intervention.

References

Brillinger, D.R. (1996). Remarks concerning graphical models for time series and point processes. *Rev. Economet.*, **16**, 1–23.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika*, **51** (2), 157–172.

Dahlhaus, R., and Eichler, M. (2003). Causality and graphical models for time series. In *Highly Structured Stochastic Systems*, Eds. P. Green, N. Hjort and S. Richardson. University Press Oxford. pp. 115-137.

Dawid, A.P. (1979). Conditional independence in statistical theory (with Discussion). *Journal of the Royal Statistical Society, Series B*, **41**, 1–31.

Dawid, A.P. (2002). Influence diagrams for causal modelling and inference. *International Statistical Review* **70**, 161–189.

Didelez, V. (2003a). Comments on graphical models for stochastic processes. In *Highly Structured Stochastic Systems*, Eds. P. Green, N. Hjort and S. Richardson. University Press Oxford.

Didelez, V. (2003b). Graphical models and sequential decisions. In *Proceedings of the 54th session of the ISI.*

Eichler, M. and Didelez, V. (2007). Causal reasoning in graphical time series models. In *Proceedings of the 23rd Annual Conference on Uncertainty in Artificial Intelligence.*

Farrow M., Goldstein M. and Spiropoulos T. (1997). Developing a Bayes linear decision support system for a brewery. In *The Practice of Bayesian Analysis* (eds S. French and J. Q. Smith), pp. 71-106. Arnold: London.

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. (1976). Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society, Series B*, **38**, 205–247.

Howard, R.A., and Matheson, J.E. (1984). Influence diagrams. In *Readings in the Principles and Applications of Decision Analysis*, Eds. R.A. Howard and J.E. Matheson. Menlo Park, CA: Strategic Decisions Group.

Kjaerulff, U. (1995) dHugin: a computational system for dynamic time-sliced Bayesian networks. *International Journal of Forecasting*, **11**, 89-111.

Lauritzen, S.L. (2000). Causal inference from graphical models. In *Complex Stochastic Systems*, Eds. O.E. Barndorff-Nielsen, D.R. Cox and C. Klüppelberg, Chapter 2, pp 63–107. London: CRC Press.

Lauritzen, S.L. and Richardson, T.S. (2002). Chain graph models and their causal interpretation (with Discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 321–361.

Oliver, R.M. and Smith, J.Q. (1990). *Influence Diagrams, Belief Nets and Decision*

Analysis. Chichester, UK: John Wiley and Sons.

Pearl, J. (1995). Causal diagrams for empirical research (with Discussion). *Biometrika*, **82**, 669–710.

Pearl, J. (2000). *Causality.* Cambridge University Press.

Queen, C.M. (1994). Using the multiregression dynamic model to forecast brand sales in a competitive product market. *The Statistician*, **43**, No. 1, 87–98.

Queen, C.M. (1997). Model elicitation in competitive markets. In *The Practice of Bayesian Analysis* (eds S. French and J. Q. Smith), pp. 229-243. London: Arnold.

Queen, C.M., Smith, J.Q. and James, D.M. (1994). Bayesian forecasts in markets with overlapping structures. *International Journal of Forecasting*, **10**, 209–233.

Queen, C.M. and Smith, J.Q. (1993). Multiregression dynamic models. *Journal of the Royal Statistical Society, Series B*, **55**, 849–870.

Queen, C.M., Wright, B.J. and Albers, C.J. (2007a). Eliciting a directed acyclic graph for a multivariate time series of vehicle counts in a traffic network. *Australian and New Zealand Journal of Statistics*, **49** (3), 221-239.

Queen, C.M., Wright, B.J. and Albers, C.J. (2007b). Forecast covariances in the linear multiregression dynamic model. *Journal of Forecasting.* (Forthcoming)

Shachter, R.D. (1986). Evaluating influence diagrams. *Operations Research*, **34**, 871–882.

Shachter, R.D. (1988). Probabilistic Inference and Influence Diagrams. *Operations Research*, **36**, 589-604.

Smith, J.Q. and Figueroa, L.J. (2007) A Causal Algebra for Dynamic Flow Networks. In *Advances in Probabilistic Graphical Models.* Springer 39–54.

Spirtes, P., Glymour, C. and Scheines, R. (2000).*Causation, Prediction and Search.* 2nd ed. New York, N.Y.: MIT Press.

Sun S.L., Zhang C.S. and Yu G.Q. (2006). A Bayesian network approach to traffic flows forecasting. *IEEE Transactions on Intelligent Transportation Systems*, **7** (1), 124–132.

West, M. and Harrison, P.J. (1986). Monitoring and adaptation in Bayesian forecasting models. *Journal of the American Statistical Association*, **81**, 741–750.

West, M. and Harrison, P.J. (1989). Subjective intervention in formal models. *Journal of Forecasting*, **8**, 33–53.

West, M. and Harrison, P.J. (1997) *Bayesian Forecasting and Dynamic Models* (2nd edition) Springer-Verlag. New York.

Whitlock, M.E. and Queen, C.M. (2000). Modelling a traffic network with missing data. *Journal of Forecasting*, **19**, 7, 561–574.