Forecast covariances in the linear multiregression dynamic model.

Short title: Forecast covariances in the linear multiregression dynamic model.

Catriona M Queen[1] **, Ben J Wright and Casper J Albers**

The Open University, Milton Keynes, MK7 6AA, UK

June 4, 2007

Abstract

The linear multiregression dynamic model (LMDM) is a Bayesian dynamic model which preserves any conditional independence and causal structure across a multivariate time series. The conditional independence structure is used to model the multivariate series by separate (conditional) univariate dynamic linear models, where each series has contemporaneous variables as regressors in its model. Calculating the forecast covariance matrix (which is required for calculating forecast variances in the LMDM) is not always straightforward in its current formulation. In this paper we introduce a simple algebraic form for calculating LMDM forecast covariances. Calculation of the covariance between model regression components can also be useful and we shall present a simple algebraic method for calculating these component covariances. In the LMDM formulation, certain pairs of series are constrained to have zero forecast covariance. We shall also introduce a possible method to relax this restriction.

Keywords: Multivariate time series, dynamic linear model, conditional independence, forecast covariance matrix, component covariances

1 Introduction.

A multiregression dynamic model (MDM) (Queen and Smith, 1993) is a non-Gaussian multivariate state space time series model. The MDM has been designed for forecasting time series for which, at any one time period, a conditional independence structure across the time series and causal drive through the system can be hypothesized. Such series can be represented by a directed acyclic graph (DAG) which, in addition to giving a useful pictorial representation of the structure of the multivariate series, is used to decompose the multivariate time series model into simpler (conditional) univariate dynamic models

[1]Correspondence to: Catriona Queen, 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

(West and Harrison, 1997). As such, the MDM is a graphical model (see, for example, Cowell *et al.* (1999)) for multivariate time series, which simplifies computation in the model through local computation. Like simultaneous equation models, in the MDM each univariate series has contemporaneous variables as regressors in its model. When these regressions are linear, we have a linear MDM (LMDM) and in this case we have a set of univariate regression dynamic linear models (DLM's) (West and Harrison, 1997, Section 9.2). Like a non time series graphical model, the conditional univariate models in an MDM are computationally simple to work with (in this case univariate DLM's), while the joint distribution can be arbitrarily complex. Also, because the model is a collection of univariate dynamic models, the MDM avoids the difficult problem of eliciting an observation covariance matrix for the multivariate problem. It also avoids the alternative of learning about the covariance matrix on-line via the matrix normal DLM (West and Harrison (1997), Section 16.4) whose use is restricted to problems in which all the individual series are similar.

Many multivariate time series may be suitable for use with an LMDM. For example, Queen (1994) uses the LMDM to model monthly brand sales in a competitive market. In this application, 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 DAG's may be elicited for market models. DAG's are also used to represent brand relationships when forecasting time series of brand sales in Goldstein *et al.* (1993), Farrow *et al.* (1997) and Farrow (2003). As another example, Whitlock and Queen (2000) and Queen *et al.* (2007) use the LMDM to model hourly vehicle counts at various points in a traffic network. Here, as in Sun *et al.* (2006), the direction of traffic flow produces the causal drive in the system and the possible routes through the network are used to define a conditional independence structure across the time series. The LMDM then produces a set of regression DLM's where contemporaneous traffic flows at upstream links in the network are used as linear regressors. Tebaldi *et al.* (2002) also use regression DLM's when modelling traffic flows. Like Queen *et al.* (2007), they use traffic flows at upstream links in the network as linear regressors. However, whereas the vehicle counts in Queen *et al.* (2007) are for one-hour intervals, those in Tebaldi *et al.* (2002) are for one-minute intervals, and so regression on lagged flows (rather than contemporaneous flows) is required. Fosen *et al.* (2006) use similar ideas to the LMDM when proposing a dynamical graphical model combining the additive hazard model and classical path analysis to analyse a trial of cancer patients with liver cirrhosis. Like the LMDM, their model uses linear regression with parents from the DAG as regressors. The focus, however, is to investigate the treatment effects on the variates over time, rather than forecasts of the variables directly.

There are many other potential application areas for the LMDM, 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).

When using the LMDM it is important to be able to calculate the forecast covariance matrix for the series. Not only is this of interest in its own right, it is also required for calculating the forecast variances for the individual series. Queen and Smith (1993) presented a recursive form for the covariance matrix, but this is not in an algebraic form which is always simple to use in practice. In this paper we introduce a simple algebraic method for calculating forecast covariances in the LMDM.

Following the superposition principle (see West and Harrison (1997), p98), each conditional univariate DLM in an LMDM can be thought of as the sum of individual model components. For example, a DLM may have a trend component, a regression component, and so on. In an LMDM it can be useful to calculate the covariance between individual DLM regression components for different series and here we introduce a simple algebraic method for calculating such component covariances.

The paper is structured as follows. In the next section the LMDM is described and its use is illustrated using a bivariate time series of brand sales. Section 3 presents a simple algebraic method for calculating the LMDM one-step and k-step ahead forecast covariance matrices. Component covariances are introduced in Section 4, along with a simple method to calculate them. In the LMDM formulation, certain pairs of series are constrained to have zero forecast covariance and Section 5 introduces a possible method to relax this restriction. Finally, Section 6 gives some concluding remarks.

2 The Linear Multiregression Dynamic Model

We have a multivariate time series $\boldsymbol{Y}_t = (Y_t(1), \ldots, Y_t(n))^T$. Suppose that the series is ordered and that the same conditional independence and causal structure is defined across the series through time, so that at each time $t = 1, 2, \ldots$, we have

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

which reads "$Y_t(r)$ is independent of $\{Y_t(1), \ldots, Y_t(r-1)\} \setminus \mathrm{pa}(Y_t(r))$ given $\mathrm{pa}(Y_t(r))$" (Dawid, 1979), where the notation "\" reads "excluding" and $\mathrm{pa}(Y_t(r)) \subseteq \{Y_t(1), \ldots, Y_t(r-1)\}$. Each variable in the set $\mathrm{pa}(Y_t(r))$ is called a *parent* of $Y_t(r)$ and $Y_t(r)$ is known as a *child* of each variable in the set $\mathrm{pa}(Y_t(r))$. We shall call any series for which $\mathrm{pa}(Y_t(\cdot)) = \varnothing$, a *root node* and list all root nodes before any children in the ordered series $\boldsymbol{Y}_t$.

The conditional independence relationships at each time point t can be represented by a DAG, where there is a directed arc to $Y_t(r)$ from each of its parents in $\mathrm{pa}(Y_t(r))$. To illustrate, Figure 1 shows a DAG for five time series at time t, where $\mathrm{pa}(Y_t(2)) = \varnothing$, $\mathrm{pa}(Y_t(3)) = \{Y_t(1), Y_t(2)\}$, $\mathrm{pa}(Y_t(4)) = \{Y_t(3)\}$ and $\mathrm{pa}(Y_t(5)) = \{Y_t(3), Y_t(4)\}$. Note that both $Y_t(1)$ and $Y_t(2)$ are root nodes.

Suppose further that there is a conditional independence and causal structure defined for the processes so that, if $\boldsymbol{Y}^t(r) = (Y_1(r), \ldots, Y_t(r))^T$,

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

Denote the information available at time t by D_t. An LMDM has the following system equation and n observation equations for all times $t = 1, 2, \ldots$.

$$\begin{array}{llcll}
\text{Observation equations:} & Y_t(r) & = & \boldsymbol{F}_t(r)^T \boldsymbol{\theta}_t(r) + v_t(r), & v_t(r) \sim N(0, V_t(r)) \quad 1 \le r \le n \\
\text{System equation:} & \boldsymbol{\theta}_t & = & G_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, & \boldsymbol{w}_t \sim N(\boldsymbol{0}, W_t) \\
\text{Initial Information:} & (\boldsymbol{\theta}_0 \,|\, D_0) & \sim & N(\boldsymbol{m}_0, C_0). &
\end{array}$$

The vector $\boldsymbol{F}_t(r)$ contains the parents $\mathrm{pa}(Y_t(r))$ and possibly other known variables (which may include $\boldsymbol{Y}^{t-1}(r)$ and $\mathrm{pa}(\boldsymbol{Y}^{t-1}(r))$); $\boldsymbol{\theta}_t(r)$ is the parameter vector for $Y_t(r)$; $V_t(1), \ldots V_t(n)$ are the scalar observation variances; $\boldsymbol{\theta}_t^T = (\boldsymbol{\theta}_t(1)^T, \ldots, \boldsymbol{\theta}_t(n)^T)$; and the matrices G_t, W_t and C_0 are all block diagonal. The error vectors, $\boldsymbol{v}_t^T = \{v_t(1), \ldots, v_t(n)\}$ and $\boldsymbol{w}_t^T = \left\{\boldsymbol{w}_t(1)^T, \ldots, \boldsymbol{w}_t(n)^T\right\}$, are such that $v_t(1), \ldots, v_t(n)$ and $\boldsymbol{w}_t(1), \ldots, \boldsymbol{w}_t(n)$ are mutually independent and $\{\boldsymbol{v}_t, \boldsymbol{w}_t\}_{t \ge 1}$ are mutually independent with time.

The LMDM therefore uses the conditional independence structure to model the multivariate time series by n separate univariate models — for $Y_t(1)$ and $Y_t(r)|\mathrm{pa}(Y_t(r))$, $r = 2, \ldots, n$. For those series with parents, each univariate model is simply a regression DLM with the parents as linear regressors. For root nodes, any suitable univariate DLM may be used. For example, consider the DAG in Figure 1. As $Y_t(1)$ and $Y_t(2)$ are both root nodes, each of these series can be modelled separately in an LMDM using any suitable univariate DLMs. $Y_t(3)$, $Y_t(4)$ and $Y_t(5)$ all have parents and so these would be modelled by (separate) univariate regression DLMs with the two regressors $Y_t(1)$ and $Y_t(2)$ for $Y_t(3)$'s model, the single regressor $Y_t(3)$ for $Y_t(4)$'s model and the two regressors $Y_t(3)$ 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 a priori, each $\boldsymbol{\theta}_t(r)$ can be updated separately in closed form from $Y_t(r)$'s (conditional) univariate model. A forecast for $Y_t(1)$ and the conditional forecasts for $Y_t(r)|\text{pa}(Y_t(r))$, $r = 2, \ldots, n$, can also be found separately using established DLM results (see West and Harrison (1997) for details). For example, in the context of the DAG in Figure 1, forecasts can be found separately (using established DLM results) for

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

However, $Y_t(r)$ and $\text{pa}(Y_t(r))$ are observed simultaneously. So the marginal forecast for each $Y_t(r)$, without conditioning on the values of its parents, is required. Unfortunately, the marginal forecast distributions for $Y_t(r)$, $r = 2, \ldots, n$, will not generally be of a simple form. However, (under quadratic loss), the mean and variance of the marginal forecast distributions for $Y_t(r)$, $r = 2, \ldots, n$, are adequate for forecasting purposes, and these can be calculated. So, returning to the context of the DAG in Figure 1, this means that the marginal forecast means and variances for $Y_t(3)$, $Y_t(4)$ and $Y_t(5)$ need to be calculated. Note that as $Y_t(1)$ and $Y_t(2)$ do not have any parents, their marginal forecasts have already been calculated from DLM theory.

For calculating the marginal forecast variance for $Y_t(r)$, the marginal forecast covariance matrix for $\text{pa}(Y_t(r))$ is required. Returning to the example DAG in Figure 1 to illustrate, this means that the marginal forecast variance of $Y_t(5)$, for example, requires the forecast covariance for $Y_t(3)$ and $Y_t(4)$, without conditioning on either of their parents $Y_t(1)$ and $Y_t(2)$ — that is, the marginal covariance of $Y_t(3)$ and $Y_t(4)$. The marginal forecast covariance matrix also provides information about the structure of the multivariate series, and, as such, is of interest in its own right. Queen and Smith (1993) gave a recursive form for calculating the marginal forecast covariance matrix. However, this is not always easy to use in practice. Section 3 presents a simple algebraic form for calculating marginal forecast covariances. First, we shall illustrate how the LMDM works in practice.

2.1 Example: forecasting a bivariate time series of brand sales

We shall illustrate using a simple LMDM for a bivariate series of monthly brand sales. The series consist of 34 months of data (supplied by Unilever Research) for two brands, which we shall call B1 and B2, who compete with each other in a product market.

Denote the sales at each time t for brands B1 and B2 by $X_t(1)$ and $X_t(2)$, respectively. Let Y_t be the total number of sales of B1 and B2 in month t (so that $Y_t = X_t(1) + X_t(2)$). Then Y_t can be modelled by a Poisson distribution with some mean μ_t, denoted $Po(\mu_t)$. Suppose that for each individual purchase of this product in month t,

$$P(\text{B1 purchased}|\text{B1 or B2 purchased}) = \theta_t.$$

Then, $X_t(1)$, the number of B1 purchased in month t, can be modelled by a binomial distribution with sample size Y_t, the total number of B1 and B2 purchased in month t, and parameter θ_t — that is, $X_t(1)|Y_t \sim Bi(Y_t, \theta_t)$.

These distributions for Y_t and $X_t(1)$ can be represented by the DAG in Figure 2. $X_t(2)$ is a logical function of its parents and is known once its parents are known. Following the terminology of WinBUGS software (http://www.mrc-bsu.cam.ac.uk/bugs/) we shall call this a logical variable and denote it on the DAG by a double oval. Note that if we had defined a conditional binomial model for $X_t(2)|Y_t$ instead, $X_t(1)$ would have been the logical variable.

Approximating the Poisson distribution for Y_t and the conditional binomial distribution for $X_t(1)|Y_t$ to normality, the observation equations in an LMDM for the DAG in Figure 2 are of the following form:

$$\begin{aligned} Y_t &= \mu_t + v_t(y), & v_t(y) &\sim N(0, V_t(y)) \\ X_t(1) &= Y_t\theta_t + v_t(1), & v_t(1) &\sim N(0, V_t(1)) \\ \text{and} \quad X_t(2) &= Y_t - X_t(1). \end{aligned}$$

The exact form of the observation equation for Y_t can be far more complicated to account for trend, seasonality and so on – whatever its form, its *mean* at time t is μ_t. The time series plot of Y_t suggests that in fact a linear growth model is appropriate, and this has been implemented here. A plot of $X_t(1)$ against Y_t is given in Figure 3. For these data, using Y_t as a linear regressor for $X_t(1)$ clearly seems sensible. If there was a non-linear relationship between Y_t and $X_t(1)$, then Y_t and/or $X_t(1)$ could be transformed that so that the linear regression model implied by the LMDM is appropriate, or the more general MDM could be used.

For simplicity, in this illustration the observation variances, $V_t(y)$ and $V_t(1)$, were fixed throughout and were simply estimated by fitting simple linear regression models for the 34 observations of Y_t and $X_t(1)$ (with regressor Y_t), respectively. Discount factors of 0.8 and 0.71 were used for Y_t and $X_t(1)$'s models, respectively, and were chosen so as to minimise the mean squared error (MSE) and the mean absolute deviation (MAD). The first two observations for each series were used to calculate initial values for the prior means for the parameters, while their initial prior variances were each set to be large (10800 for the two parameters for Y_t and 1 for the parameter for $X_t(1)$) to allow forecasts to adapt quickly.

To get an idea of how well the LMDM is performing for these series, a standard multivariate DLM (MV DLM) was also used, using linear growth models for both $X_t(1)$ and $X_t(2)$. In order to try to get a fair comparison, the observation covariance matrix was assumed fixed and was calculated in exactly the same way as the observation variances were for the LMDM. As with the LMDM, initial estimates of prior means for the parameters were calculated using the first two observations of the series, the associated prior variances were set to be large to allow quick adaptation in the model and discount factors were chosen (0.85 for $X_t(1)$, 0.70 for $X_t(2)$) so as to minimise the MSE/MAD.

Figure 4 shows plots of the time series for $X_t(1)$ and $X_t(2)$, together with the (marginal) one-step forecasts and ± 1.96 (marginal) forecast standard deviation error bars calculated using both the MV DLM and the LMDM. From these plots, it can be seen that the LMDM is performing better than the MV DLM. This is also clearly reflected in the MSE/MAD values for the one-step ahead forecasts for the two models, shown in Table 1. Also shown in Table 1 are the MSE/MAD values for the two- and three-step ahead forecasts using both models. Clearly, the LMDM is also performing better with respect to these k-step ahead forecasts.

In addition to giving a better forecast performance, the LMDM also has the advantage that intervention can be simpler to implement. For example, suppose that total sales of the brands were expected to increase suddenly. Then to accommodate this information into the model, only intervention for Y_t would be required in the LMDM, whereas the MV DLM would require intervention for both $X_t(1)$ and $X_t(2)$. The LMDM also has the advantage that only observation variances need to be elicited, whereas the MV DLM has the additional difficult task of eliciting the observation covariance between $X_t(1)$ and

$X_t(2)$ as well.

3 Simple calculation of marginal forecast covariances

From Queen and Smith (1993), the marginal forecast covariance between $Y_t(i)$ and $Y_t(r)$, $i < r$, $r = 2, \ldots, n$, can be calculated recursively using,

$$\begin{aligned} \mathrm{cov}(Y_t(i), Y_t(r)|D_{t-1}) &= \mathrm{E}\left(Y_t(i) \cdot \mathrm{E}\left(Y_t(r)|Y_t(1), \ldots, Y_t(r-1), D_{t-1}\right)|D_{t-1}\right) \\ &\quad -\mathrm{E}(Y_t(i)|D_{t-1})\mathrm{E}(Y_t(r)|D_{t-1}). \end{aligned} \tag{3.1}$$

In this paper we shall use this to derive a simple algebraic form for calculating these forecast covariances. In what follows let $\boldsymbol{a}_t(r)$ be the prior mean for $\boldsymbol{\theta}_t(r)$.

Theorem 1 *In an LMDM, let $Y_t(j_1), \ldots, Y_t(j_{m_r})$ be the m_r parents of $Y_t(r)$. Then for $i < r$, the one-step ahead forecast covariance between $Y_t(i)$ and $Y_t(r)$ can be calculated using*

$$cov(Y_t(i), Y_t(r)|D_{t-1}) = \sum_{l=1}^{m_r} cov(Y_t(i), Y_t(j_l)|D_{t-1}) a_t(r)^{(j_l)},$$

where $a_t(r)^{(j_l)}$ is the element of $\boldsymbol{a}_t(r)$ associated with the parent regressor $Y_t(j_l)$ — ie $a_t(r)^{(j_l)}$ is the prior mean for the parameter for regressor $Y_t(j_l)$.

Proof. Consider Equation 3.1. From the observation equations for the LMDM,

$$\mathrm{E}\left(Y_t(r)|Y_t(1), \ldots, Y_t(r-1), D_{t-1}\right) = \boldsymbol{F}_t(r)^T \boldsymbol{a}_t(r).$$

Also, using the result that for two random variables X and Y, $\mathrm{E}(Y) = \mathrm{E}(\mathrm{E}(Y|X))$,

$$\mathrm{E}\left(Y_t(r)|D_{t-1}\right) = \mathrm{E}\left(\boldsymbol{F}_t(r)^T \boldsymbol{a}_t(r)|D_{t-1}\right) = \mathrm{E}\left(\boldsymbol{F}_t(r)^T|D_{t-1}\right) \boldsymbol{a}_t(r).$$

So

$$\begin{aligned} \mathrm{cov}(Y_t(i), Y_t(r)|D_{t-1}) &= \mathrm{E}(Y_t(i) \cdot \boldsymbol{F}_t(r)^T|D_{t-1})\boldsymbol{a}_t(r) - \mathrm{E}(Y_t(i)|D_{t-1})\mathrm{E}(\boldsymbol{F}_t(r)^T|D_{t-1})\boldsymbol{a}_t(r) \\ &= \mathrm{cov}(Y_t(i), \boldsymbol{F}_t(r)^T|D_{t-1})\boldsymbol{a}_t(r). \end{aligned}$$

Now $Y_t(r)$ has the m_r parents $Y_t(j_1), \ldots, Y_t(j_{m_r})$, so

$$\boldsymbol{F}_t(r)^T = \begin{pmatrix} Y_t(j_1) & \cdots & Y_t(j_{m_r}) & \boldsymbol{x}_t(r)^T \end{pmatrix}$$

where $\boldsymbol{x}_t(r)^T$ is a vector of known variables. Then, $\mathrm{cov}(Y_t(i), \boldsymbol{x}_t(r)^T|D_{t-1})$ is simply a vector of zeros and so

$$\mathrm{cov}(Y_t(i), Y_t(r)|D_{t-1}) = \sum_{l=1}^{m_r} \mathrm{cov}(Y_t(i), Y_t(j_l)|D_{t-1}) a_t(r)^{(j_l)}$$

as required. ■

The marginal forecast covariance between $Y_t(i)$ and $Y_t(r)$ is therefore simply the sum of the covariances between $Y_t(i)$ and each of the parents of $Y_t(r)$. Consequently it is simple to calculate the one-step forecast covariance matrix recursively.

Corollary 1 *In an LMDM, let $Y_t(j_1), \ldots, Y_t(j_{m_r})$ be the m_r parents of $Y_t(r)$. Suppose that the series have been observed up to, and including, time t. Define*

$$\boldsymbol{a}_t(r,k) = E(\boldsymbol{\theta}_{t+k}(r)|D_t)$$

so that

$$\boldsymbol{a}_t(r,k) = G_{t+k}(r)\boldsymbol{a}_t(r,k-1)$$

with $\boldsymbol{a}_t(r,0) = \boldsymbol{m}_t(r)$, the posterior mean for $\boldsymbol{\theta}_t(r)$ at time t. Denote the parameter associated with parent regressor $Y_t(j_l)$ by $\theta_t(r)^{(j_l)}$, and $E(\theta_{t+k}(r)^{(j_l)}|D_t)$ by $a_t(r,k)^{(j_l)}$, the associated element of $\boldsymbol{a}_t(r,k)$.

Then, for $i < r$ and $k \geq 1$, the k-step ahead forecast covariance between $Y_{t+k}(i)$ and $Y_{t+k}(r)$ can be calculating recursively using

$$cov(Y_{t+k}(i), Y_{t+k}(r)|D_t) = \sum_{l=1}^{m_r} cov(Y_{t+k}(i), Y_{t+k}(j_l)|D_t) a_t(r,k)^{(j_l)}.$$

Proof. Let $\boldsymbol{X}_{t+k}(r)^T = \begin{pmatrix} Y_{t+k}(1) & \ldots & Y_{t+k}(r-1) \end{pmatrix}$. Using the result that $E(XY) = E[X \cdot E(Y|X)]$,

$$\begin{aligned} \text{cov}\left(\boldsymbol{X}_{t+k}(r), Y_{t+k}(r)|D_t\right) &= E(\boldsymbol{X}_{t+k}(r) \cdot E(Y_{t+k}(r)|Y_{t+k}(1), \ldots, Y_{t+k}(r-1), D_t)|D_t) \\ &\quad - E(\boldsymbol{X}_{t+k}(r)|D_t)E(Y_{t+k}(r)|D_t). \end{aligned} \tag{3.2}$$

From DLM theory (see, for example, West & Harrison (1997), pp 106–7),

$$E(Y_{t+k}(r)|Y_{t+k}(1), \ldots, Y_{t+k}(r-1), D_t) = \boldsymbol{F}_{t+k}(r)^T \boldsymbol{a}_t(r,k) \tag{3.3}$$

where

$$\boldsymbol{a}_t(r,k) = G_{t+k}(r)\boldsymbol{a}_t(r,k-1)$$

with $\boldsymbol{a}_t(r,0) = \boldsymbol{m}_t(r)$, the posterior mean for $\boldsymbol{\theta}_t(r)$ at time t. Also,

$$\begin{aligned} E(Y_{t+k}(r)|D_t) &= E[E(Y_{t+k}(r)|Y_{t+k}(1), \ldots, Y_{t+k}(r-1), D_t)] \\ &= E[\boldsymbol{F}_{t+k}(r)^T \boldsymbol{a}_t(r,k)|D_t] \qquad \text{by Equation 3.3.} \end{aligned}$$

Thus Equation 3.2 becomes,

$$\text{cov}(\boldsymbol{X}_{t+k}(r), Y_{t+k}(r)|D_t) = \text{cov}(\boldsymbol{X}_{t+k}(r), \boldsymbol{F}_{t+k}(r)^T|D_t)\boldsymbol{a}_t(r,k)$$

and so the ith row of this gives us

$$\text{cov}(Y_{t+k}(i), Y_{t+k}(r)|D_t) = \sum_{l=1}^{m_r} \text{cov}(Y_{t+k}(i), Y_{t+k}(j_l)|D_t) a_t(r,k)^{(j_l)},$$

as required. ■

Corollary 2 *For two root nodes $Y_t(i)$ and $Y_t(r)$, under the LMDM,*

$$cov(Y_t(i), Y_t(r)|D_{t-1}) = 0.$$

Proof. From the proof of Theorem 1,

$$\text{cov}(Y_t(i), Y_t(r)|D_{t-1}) = \text{cov}(Y_t(i), \boldsymbol{F}_t(r)^T|D_{t-1})\boldsymbol{a}_t(r).$$

Since $Y_t(r)$ is a root node, $\boldsymbol{F}_t(r)$ only contains known variables so that,

$$\text{cov}(Y_t(i), \boldsymbol{F}_t(r)^T|D_{t-1}) = 0.$$

The result then follows directly. ■

To illustrate calculating a one-step forecast covariance using Theorem 1 and Corollary 2, consider the following example.

Example 1 Consider the DAG in Figure 1. For $r = 2, \ldots, 5$ and $j = 1, \ldots, 4$, let $a_t(r)^{(j)}$ be the prior mean for parent regressor $Y_t(j)$ in $Y_t(r)$'s model. The forecast covariance between $Y_t(1)$ and $Y_t(5)$, for example, is then calculated as follows.

$$\begin{aligned}
\text{cov}(Y_t(1), Y_t(5)|D_{t-1}) &= \text{cov}(Y_t(1), Y_t(4)|D_{t-1})a_t(5)^{(4)} + \text{cov}(Y_t(1), Y_t(3)|D_{t-1})a_t(5)^{(3)} \\
\text{cov}(Y_t(1), Y_t(4)|D_{t-1}) &= \text{cov}(Y_t(1), Y_t(3)|D_{t-1})a_t(4)^{(3)} \\
\text{cov}(Y_t(1), Y_t(3)|D_{t-1}) &= \text{cov}(Y_t(1), Y_t(1)|D_{t-1})a_t(3)^{(1)} + \text{cov}(Y_t(1), Y_t(2)|D_{t-1})a_t(3)^{(2)}.
\end{aligned}$$

$Y_t(1)$ and $Y_t(2)$ are both root nodes and so their covariance is 0. So,

$$\text{cov}(Y_t(1), Y_t(5)|D_{t-1}) = \text{var}(Y_t(1)|D_{t-1})a_t(3)^{(1)} \left(a_t(4)^{(3)} a_t(5)^{(4)} + a_t(5)^{(3)} \right).$$

where $\text{var}(Y_t(1)|D_{t-1})$ is the marginal forecast variance for $Y_t(1)$. This is both simple and fast to calculate. ■

When a DAG has a logical node, the logical node will be some (logical) function of its parents. If the function is linear, then the covariance between $Y_t(i)$ and a logical node $Y_t(r)$ will simply be a linear function of the covariance between $Y_t(i)$ and the parents of $Y_t(r)$. For example, for the DAG in Figure 5, suppose that $Y_t(4) = Y_t(3) - Y_t(2)$. Then

$$\mathrm{cov}(Y_t(1), Y_t(4)|D_{t-1}) = \mathrm{cov}(Y_t(1), Y_t(3)|D_{t-1}) - \mathrm{cov}(Y_t(1), Y_t(2)|D_{t-1}).$$

The covariance can then be found simply by applying Theorem 1.

4 Component covariances

Suppose that $Y_t(r)$ has the m_r parents $Y_t(r'_1), \ldots, Y_t(r'_{m_r})$ for $r = 2, \ldots, n$. Write the observation equation for each $Y_t(r)$ as the sum of regression components as follows.

$$Y_t(r) = \sum_{l=1}^{m_r} Y_t(r, r'_l) + Y_t(r, \boldsymbol{x}_t(r)) + v_t(r), \qquad v_t(r) \sim N(0, V_t(r)) \tag{4.1}$$

with

$$\begin{aligned} Y_t(r, r'_l) &= Y_t(r'_l)\theta_t(r)^{(r'_l)} \\ Y_t(r, \boldsymbol{x}_t(r)) &= \boldsymbol{x}_t(r)^T \boldsymbol{\theta}_t(r)^{(\boldsymbol{x}_t(r))} \end{aligned}$$

where $\theta_t(r)^{(r'_l)}$ is the parameter associated with the parent regressor $Y_t(r'_l)$ and $\boldsymbol{\theta}_t(r)^{(\boldsymbol{x}_t(r))}$ is the vector of parameters for known variables $\boldsymbol{x}_t(r)$. It can sometimes be helpful to find the covariance between two components $Y_t(i, i')$ and $Y_t(r, r')$ from the models for $Y_t(i)$ and $Y_t(r)$ respectively. Call $\mathrm{cov}(Y_t(i, i'), Y_t(r, r'))$ the *component covariance* for $Y_t(i, i')$ and $Y_t(r, r')$. We shall illustrate why component covariances may be useful by considering two examples.

Example 2 Component covariances in traffic networks

Queen *et al.* (2007) consider the problem of forecasting hourly vehicle counts at various points in a traffic network. The possible routes through the network are used to elicit a DAG for use with an LMDM. It may be useful to learn about driver route choice probabilities in such a network — i.e. the probability that a vehicle starting at a certain point A will travel to destination B. Unfortunately, these are not always easy to estimate from vehicle count data. However, component covariances could be useful in this respect, as illustrated by the following hypothetical example.

Consider the simple traffic network illustrated in Figure 6. There are five data collection sites, each of which records the hourly count of vehicles passing that site. There are four possible routes through the system: A to C, A to D, B to C and B to D. Because all traffic from A and B to C and D is counted at site 3, it can be difficult to learn about driver route choices using the time series of vehicle counts alone.

Let $Y_t(r)$ be the vehicle count for hour t at site r. A suitable DAG representing the conditional independence relationships between $Y_t(1), \ldots, Y_t(5)$ is given in Figure 7. (For details on how this DAG can be elicited see Queen *et al.* (2007).) Notice that all vehicles at site 3 flow to sites 4 or 5 so that conditional on $Y_t(3)$ and $Y_t(4)$, $Y_t(5)$ is a logical node with $Y_t(5) = Y_t(3) - Y_t(4)$.

From Figure 6, $Y_t(3)$ receives all its traffic from sites 1 and 2, while $Y_t(4)$ receives all its traffic from site 3. So possible LMDM observation equations for $Y_t(3)$ and $Y_t(4)$ are given by,

$$\begin{aligned}
Y_t(3) &= Y_t(1)\theta_t(3)^{(1)} + Y_t(2)\theta_t(3)^{(2)} + v_t(3), & v_t(3) &\sim N(0, V_t(3)), \\
Y_t(4) &= Y_t(3)\theta_t(4)^{(3)} + v_t(4), & v_t(4) &\sim N(0, V_t(4)).
\end{aligned}$$

where $0 \leq \theta_t(3)^{(1)}, \theta_t(3)^{(2)}, \theta_t(4)^{(3)} \leq 1$. Then

$$\begin{aligned}
Y_t(1)\theta_t(3)^{(1)} &= Y_t(3,1) = \text{number of vehicles travelling from site 1 to 3 in hour } t, \\
Y_t(3)\theta_t(4)^{(3)} &= Y_t(4,3) = \text{number of vehicles travelling from site 3 to 4 in hour } t.
\end{aligned}$$

So $\text{cov}(Y_t(3,1), Y_t(4,3))$ is informative about the use of route A to C. A high correlation between $Y_t(3,1)$ and $Y_t(4,3)$ indicates a high probability that a vehicle at A travels to C and a small correlation indicates a low probability. Of course, the actual driver route choice probabilities still cannot be estimated from these data. However having some idea of the relative magnitude of the choice probabilities can still be very useful. ■

Example 3 Accommodating changes in the DAG

In many application areas the DAG representing the multivariate time series may change over time — either temporarily or permanently. For example, in a traffic network a temporary diversion may change the DAG temporarily, or a change in the road layout may change the DAG permanently. Because of the structure of the LMDM, much of the posterior information from the original DAG can be used to help form priors in the new DAG (see Queen *et al.* (2007) for an example illustrating this). In this respect, component covariances may be informative about covariances in the new DAG. We shall illustrate this using a simple example.

Figure 8 shows part of a DAG representing four time series $Y_t(1), \ldots, Y_t(4)$ (the DAG continues with children of $Y_t(3)$ and $Y_t(4)$, but we are only interested in $Y_t(1), \ldots, Y_t(4)$ here). Using an LMDM and Equation 4.1, suppose we have the following observation equations for $Y_t(3)$ and $Y_t(4)$:

$$Y_t(3) = Y_t(3,1) + v_t(3), \qquad v_t(3) \sim N(0, V_t(3)), \tag{4.2}$$

$$Y_t(4) = Y_t(4,1) + Y_t(4,2) + v_t(4), \qquad v_t(4) \sim N(0, V_t(4)). \tag{4.3}$$

Now suppose that the DAG is changed so that a new series, X_t, is introduced which lies between $Y_t(1)$ and $Y_t(4)$ in Figure 8. Suppose further that $Y_t(1)$ and $Y_t(4)$ are no longer observed. The new DAG is given in Figure 9 (again the DAG continues, this time with children of $Y_t(3)$, X_t and $Y_t(2)$). The component $Y_t(3,1)$ from Equation 4.2 is informative about $Y_t(3)$ in the new model, and the component $Y_t(4,1)$ from Equation 4.3 is informative about X_t in the new model. Thus the component covariance $\text{cov}(Y_t(3,1), Y_t(4,1))$ is informative about $\text{cov}(Y_t(3), X_t)$ in the new model. ■

The following theorem presents a simple method for calculating component covariances.

Theorem 2 *Suppose that $Y_t(i')$ is the parent of $Y_t(i)$ and $Y_t(r')$ is a parent of $Y_t(r)$, with $i < r$. Then,*

$$cov(Y_t(i,i'), Y_t(r,r')|D_{t-1}) = cov(Y_t(i'), Y_t(r')|D_{t-1}) a_t(i)^{(i')} a_t(r)^{(r')}$$

where $a_t(i)^{(i')}$ and $a_t(r)^{(r')}$ are, respectively, the prior means for the regressor $Y_t(i')$ in $Y_t(i)$'s model and the regressor $Y_t(r')$ in $Y_t(r)$'s model.

Proof. For each r, let

$$\boldsymbol{Z}_t^{(r)^T} = \begin{pmatrix} Y_t(r,r'_1) & Y_t(r,r'_2) & \ldots & Y_t(r,r'_{m_r}) & Y_t(r,\boldsymbol{x}_t(r)) & v_t(r) \end{pmatrix}$$

and

$$\boldsymbol{Z}_t(r)^T = \begin{pmatrix} \boldsymbol{Z}_t^{(1)^T} & \boldsymbol{Z}_t^{(2)^T} & \ldots & \boldsymbol{Z}_t^{(r-1)^T} \end{pmatrix}.$$

Using the result that for two random variables X and Y, $\text{E}(XY) = \text{E}(X \cdot \text{E}(Y|X))$, we have, for a specific $r' \in \{r'_1, \ldots, r'_{m_r}\}$,

$$\begin{aligned} \text{cov}(\boldsymbol{Z}_t(r), Y_t(r,r')|D_{t-1}) &= \text{E}(\boldsymbol{Z}_t(r) \cdot \text{E}(Y_t(r,r')|\boldsymbol{Z}_t(r), D_{t-1})|D_{t-1}) \\ &\quad - \text{E}(\boldsymbol{Z}_t(r)|D_{t-1})\text{E}(Y_t(r,r')|D_{t-1}). \end{aligned} \tag{4.4}$$

Now, from Equation 4.1,

$$\mathrm{E}(Y_t(r,r')|\boldsymbol{Z}_t(r),D_{t-1}) = \mathrm{E}(Y_t(r,r')|Y_t(1),\ldots,Y_t(r-1),D_{t-1}) = Y_t(r')a_t(r)^{(r')}.$$

Also,

$$\mathrm{E}(Y_t(r,r')|D_{t-1}) = \mathrm{E}[\mathrm{E}(Y_t(r,r')|\boldsymbol{Z}_t(r),D_{t-1})] = \mathrm{E}(Y_t(r')a_t(r)^{(r')}|D_{t-1}).$$

So, Equation 4.4 becomes,

$$\begin{aligned}\mathrm{cov}(\boldsymbol{Z}_t(r),Y_t(r,r')|D_{t-1}) &= \mathrm{E}(\boldsymbol{Z}_t(r)Y_t(r')a_t(r)^{(r')}|D_{t-1}) \\ &\quad -\mathrm{E}(\boldsymbol{Z}_t(r)|D_{t-1})\mathrm{E}(Y_t(r')a_t(r)^{(r')}|D_{t-1}) \\ &= \mathrm{cov}(\boldsymbol{Z}_t(r),Y_t(r')|D_{t-1})a_t(r)^{(r')}.\end{aligned}$$

The single row from $\boldsymbol{Z}_t(r)$ corresponding to $Y_t(i,i')$ then gives us,

$$\mathrm{cov}(Y_t(i,i'),Y_t(r,r')|D_{t-1}) = \mathrm{cov}(Y_t(i,i'),Y_t(r')|D_{t-1})a_t(r)^{(r')}. \tag{4.5}$$

Let

$$\boldsymbol{X}_t(r,i)^T = \begin{pmatrix} Y_t(1) & Y_t(2) & \ldots & Y_t(i-1) & Y_t(i+1) & \ldots & Y_t(r-1) \end{pmatrix}.$$

Then

$$\begin{aligned}\mathrm{cov}(\boldsymbol{X}_t(r,i),Y_t(i,i')|D_{t-1}) &= \mathrm{E}(\boldsymbol{X}_t(r,i)\cdot\mathrm{E}(Y_t(i,i')|\boldsymbol{X}_t(r,i),D_{t-1})|D_{t-1}) \\ &\quad -\mathrm{E}(\boldsymbol{X}_t(r,i)|D_{t-1})\mathrm{E}(Y_t(i,i')|D_{t-1}). \end{aligned}\tag{4.6}$$

Now,

$$\mathrm{E}(Y_t(i,i')|\boldsymbol{X}_t(r,i),D_{t-1}) = Y_t(i')a_t(i)^{(i')},$$

and Equation 4.6 becomes

$$\mathrm{cov}(\boldsymbol{X}_t(r,i),Y_t(i,i')|D_{t-1}) = \mathrm{cov}(\boldsymbol{X}_t(r,i),Y_t(i')|D_{t-1})a_t(i)^{(i')}.$$

So taking the individual row of $\boldsymbol{X}_t(r,i)$ corresponding to $Y_t(r')$ we get,

$$\mathrm{cov}(Y_t(r'),Y_t(i,i')|D_{t-1}) = \mathrm{cov}(Y_t(r'),Y_t(i')|D_{t-1})a_t(i)^{(i')}.$$

Substituting this into Equation 4.5 gives us

$$\mathrm{cov}(Y_t(i,i'),Y_t(r,r')|D_{t-1}) = \mathrm{cov}(Y_t(i'),Y_t(r')|D_{t-1})a_t(i)^{(i')}a_t(r)^{(r')}$$

as required. ■

Theorem 2 allows the simple calculation of the component correlations. For example, in Example 2,

$$\mathrm{cov}(Y_t(3,1), Y_t(4,3)|D_{t-1}) = \mathrm{cov}(Y_t(1), Y_t(3)|D_{t-1})a_t(3)^{(1)}a_t(4)^{(3)}$$

and $\mathrm{cov}(Y_t(1), Y_t(3)|D_{t-1})$ is simple to calculate using Theorem 1.

Corollary 3 *Suppose that $Y_t(i')$ is the parent of $Y_t(i)$ and $Y_t(r')$ is a parent of $Y_t(r)$, with $i < r$. Then, using the notation presented in Corollary 1, for $i < r$ and $k \geq 1$, the k-step ahead forecast covariance between components $Y_{t+k}(i,i')$ and $Y_{t+k}(r,r')$ can be calculating recursively using*

$$cov(Y_{t+k}(i,i'), Y_{t+k}(r,r')|D_t) = \sum_{l=1}^{m_r} cov(Y_{t+k}(i'), Y_{t+k}(r')|D_t)a_t(i,k)^{(i')}a_t(r,k)^{(r')}.$$

Proof. This is proved using the same argument to that used in the proof of Theorem 2, where t is replaced by $t+k$ and noting the $\mathrm{E}(\theta_{t+k}(i)^{(i')}|D_t) = a_t(i,k)^{(i')}$ and $\mathrm{E}(\theta_{t+k}(r)^{(r')}|D_t) = a_t(r,k)^{(r')}$. ■

5 Covariance between root nodes

Recall from Corollary 2 that the covariance between two root nodes is zero in the LMDM. However, this is not always appropriate. For example, consider the traffic network in Example 2. Both $Y_t(1)$ and $Y_t(2)$ are root nodes which may in fact be highly correlated — they may have similar daily patterns with the same peak times, etc, and they may be affected in a similar way by external events such as weather conditions.

One possible way to introduce non-zero covariances between root nodes is to add an extra node as a parent of all root nodes in the DAG. This extra node represents any variables which may account for the correlation between the root nodes.

Example 4 In the traffic network in Example 2, suppose that $\boldsymbol{X}_t$ is a vector of variables which can account for the correlation between $Y_t(1)$ and $Y_t(2)$. So, $\boldsymbol{X}_t$ might include such variables as total traffic volume entering the system, hourly rainfall, temperature, and so on. Then $\boldsymbol{X}_t$ can be introduced into the DAG as a parent of $Y_t(1)$ and $Y_t(2)$ as in Figure 10.

The observation equations for $Y_t(1)$ and $Y_t(2)$ now both have $\boldsymbol{X}_t$ as regressors and applying Theorem 1 we get,

$$\text{cov}(Y_t(1), Y_t(2)|D_{t-1}) = \text{var}(\boldsymbol{X}_t|D_{t-1})\boldsymbol{a}_t(1)^{(\boldsymbol{X}_t)}\boldsymbol{a}_t(2)^{(\boldsymbol{X}_t)}$$

where $\boldsymbol{a}_t(2)^{(\boldsymbol{X}_t)}$ and $\boldsymbol{a}_t(2)^{(\boldsymbol{X}_t)}$ are the prior mean vectors for regressors $\boldsymbol{X}_t$ in $Y_t(1)$ and $Y_t(2)$'s model. ■

6 Concluding remarks

In this paper we have presented a simple algebraic form for calculating the one step ahead covariance matrix in LMDMs. We have also introduced a simple method for calculating covariances between regression components of different DLMs within the LMDM. Component covariances may be of interest in their own right, and may also prove to be useful for forming informative priors following any changes in the DAG for the LMDM. Their use in practice now needs to be investigated in further research.

One of the problems with the LMDM is the imposition of zero covariance between root nodes. To allow nonzero covariance between root nodes we have proposed introducing $\boldsymbol{X}_t$ into the model as a parent of all the root nodes, where $\boldsymbol{X}_t$ is a set of variables which may help to explain the correlation between parents. Further research is now required to investigate how well this might work in practice.

Acknowledgements

We would like to thank the reviewer for their very useful comments on an earlier version of this paper and also Unilever Research for supplying the data.

References

Cowell RG, Dawid AP, Lauritzen SL and Spiegelhalter DJ. 1999. *Probabilistic Networks and Expert Systems.* Springer Verlag: New York.

Dawid AP. 1979. Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society*, **41**, 1–31.

Farrow M. 2003. Practical building of subjective covariance structures for large complicated systems. *The Statistician*, **52**, No. 4, 553–573.

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.

Fosen J, Ferkingstad E, Borgan ∅, and AalenFosen OO. 2006. Dynamic path analysisa new approach to analyzing time-dependent covariates. *Lifetime Data Analysis*, **12** (2), 143–167.

Goldstein M, Farrow M and Spiropoulos T. 1993. Prediction under the influence: Bayes linear influence diagrams for prediction in a large brewery. *The Statistician*, **42**, No. 4, 445-459.

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

Queen CM. 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 CM, Smith JQ and James DM. 1994. Bayesian forecasts in markets with overlapping structures. *International Journal of Forecasting*, **10**, 209–233.

Queen CM and Smith JQ. 1993. Multiregression dynamic models. *Journal of the Royal Statistical Society, Series B*, **55**, 849–870.

Queen CM, Wright BJ and Albers CJ. 2007. Eliciting a directed acyclic graph for a multivariate time series of vehicle counts in a traffic network. *Australian and New Zealand Journal of Statistics* To appear.

Sun SL, Zhang CS and Yu GQ. 2006. A Bayesian network approach to traffic flows forecasting. *IEEE Transactions on Intelligent Transportation Systems*, **7** (1), 124–132.

Tebaldi C, West M and Karr AK. 2002. Statistical analyses of freeway traffic flows. *Journal of Forecasting*, **21**, 39-68.

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

Whitlock ME and Queen CM. 2000. Modelling a traffic network with missing data. *Journal of Forecasting*, **19**, 7, 561–574.

Authors' biographies

Catriona Queen is a Lecturer in Statistics in the Statistics Department at The Open University, UK. Her research interests include dynamic models, multivariate forecasting, graphical models and Bayesian Statistics.

Casper Albers is a Research Fellow in the Statistics Department at The Open University, UK. His research interests include multivariate forecasting, Bayesian statistics and statistical geometry.

Ben Wright Benjamin Wright completed his PhD in 2005 in the Statistics Department at the Open University, UK, and his research interests include Bayesian statistics, dynamic models, multivariate time series and Markov chain Monte Carlo methods.

Authors' addresses

Catriona Queen, Department of Statistics, The Open University, Milton Keynes, MK7 6AA, UK.

Casper Albers, Department of Statistics, The Open University, Milton Keynes, MK7 6AA, UK.

Ben Wright, Department of Statistics, The Open University, Milton Keynes, MK7 6AA, UK.

		MSE		MAD	
		MV DLM	LMDM	MV DLM	LMDM
One-step forecasts	$X_t(1)$	3541.72	1601.39	26.89	23.26
	$X_t(2)$	4622.96	3988.42	49.31	40.77
Two-step forecasts	$X_t(1)$	4579.25	1513.81	27.71	23.56
	$X_t(2)$	5486.01	3998.98	51.96	37.86
Three-step forecasts	$X_t(1)$	3854.69	1509.43	39.54	22.51
	$X_t(2)$	7198.47	4145.12	68.43	39.56

Table 1: MSE and MAD values for the MV DLM and LMDM.

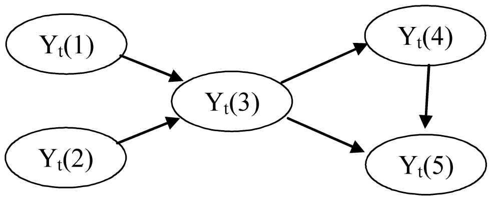

Figure 1: DAG representing five time series at time t.

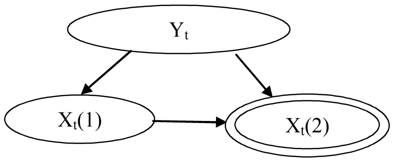

Figure 2: DAG representing monthly sales of brands B1 and B2 ($X_t(1)$ and $X_t(2)$, respectively) and their sum (Y_t).

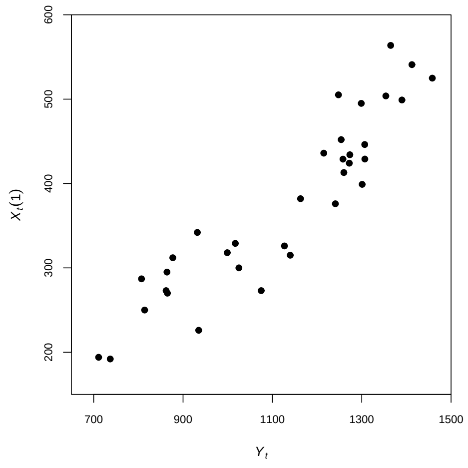

Figure 3: Plot of $X_t(1)$ against Y_t.

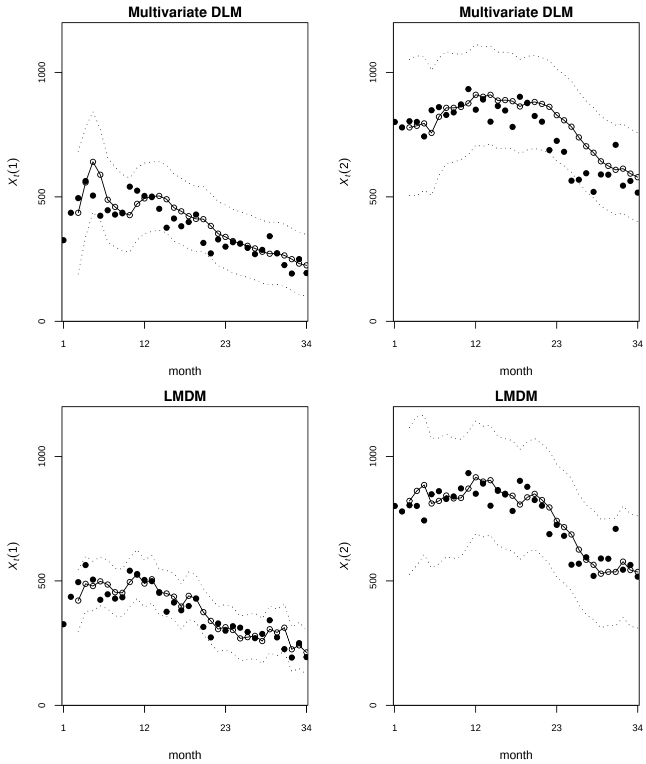

Figure 4: Plots of the one-step forecasts (joined dots) and ±1.96 forecast standard deviation error bars (dotted lines) for $X_t(1)$ and $X_t(2)$ calculated using the LMDM, together with the data (filled dots).

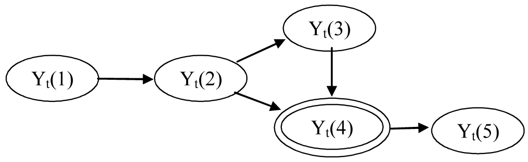

Figure 5: DAG with a logical node $Y_t(4)$.

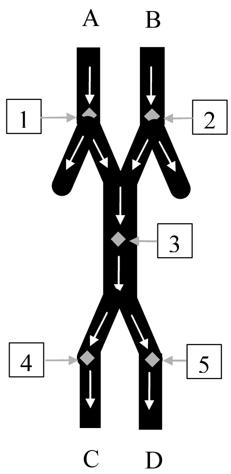

Figure 6: Hypothetical traffic network for Example 2 with starting points A and B and destinations C and D. The grey diamonds are the data collection sites, each of which is numbered. White arrows indicate the direction of traffic flow.

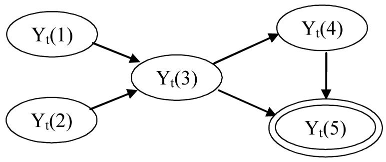

Figure 7: DAG representing the conditional independence relationships between $Y_t(1), \ldots, Y_t(5)$ in the traffic network of Figure 6.

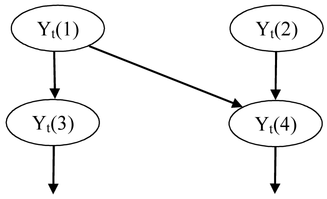

Figure 8: Part of a DAG for Example 3 for four time series $Y_t(1), \ldots, Y_t(4)$.

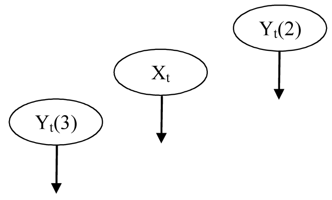

Figure 9: New DAG for Example 3 (adapting the DAG in Figure 8) where X_t is introduced into the DAG and $Y_t(1)$ and $Y_t(4)$ are no longer observed.

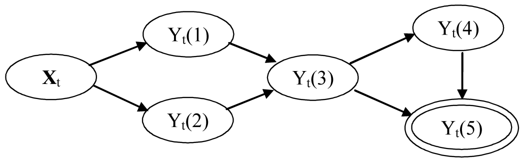

Figure 10: DAG representing time series of vehicle counts in Example 4, with an extra node $\boldsymbol{X}_t$ to account for any correlation between $Y_t(1)$ and $Y_t(2)$.