Covariance calculations in the linear multiregression dynamic model.

Ben J Wright (The Open University)

Catriona M Queen (The Open University)[1]

July 21, 2005

Summary

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 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 linear multiregression dynamic model (LMDM) (Queen and Smith, 1993) is a multivariate Bayesian dynamic model (West and Harrison, 1997). The LMDM is suitable for any multivariate time series for which there is a conditional independence structure and some causal driving mechanism within the system. 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. As another example, Queen and Wright (2005) use the LMDM to model hourly vehicle counts at various points in a traffic network. Here, the direction of traffic flow produces the causal drive in the system and the possible

[1]Address for correspondence: Department of Statistics, Faculty of Mathematics and Computing, The Open University, Milton Keynes, MK7 6AA, UK. Tel: (+44) 01908 659585 Fax: (+44) 01908 652140 Email: C.Queen@open.ac.uk

routes through the network are used to define a conditional independence structure across the time series. There are many other potential application areas, 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), medical problems (such as patient physiological monitoring), and so on.

The LMDM uses the conditional independence and causal structure within the system to decompose a multivariate time series model into simpler (conditional) univariate dynamic linear models (DLMs) (West and Harrison, 1997), where each series has contemporaneous variables as regressors in its model. The LMDM is analogous to a non time series graphical model in that the conditional univariate models are computationally simple to work with (in this case univariate DLMs), while the joint distribution can be highly complex.

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.

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 then Section 3 presents a simple algebraic method for calculating the LMDM one step ahead forecast covariance matrix. 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.

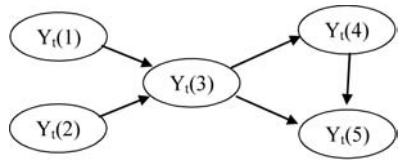

Figure 1: DAG representing five time series at time t, where $\text{pa}(Y_t(2)) = \varnothing$, $\text{pa}(Y_t(3)) = \{Y_t(1), Y_t(2)\}$, $\text{pa}(Y_t(4)) = \{Y_t(3)\}$ and $\text{pa}(Y_t(5)) = \{Y_t(3), Y_t(4)\}$. Both $Y_t(1)$ and $Y_t(2)$ are root nodes.

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

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

which reads "$Y_t(r)$ is independent of $\{Y_t(1), \ldots, Y_t(r-1)\} \backslash \text{pa}(Y_t(r))$ given $\text{pa}(Y_t(r))$" (using the notation that "\" reads "excluding"), where $\text{pa}(Y_t(r)) \subseteq \{Y_t(1), \ldots, Y_t(r-1)\}$. Suppose further that there is a conditional independence and causal structure defined for the processes so that, if $\boldsymbol{Y}^t(r) = (Y_1(r), \ldots, Y_t(r))^T$,

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

Each variable in the set $\text{pa}(Y_t(r))$ is called a *parent* of $Y_t(r)$ and $Y_t(r)$ is known as a *child* of each variable in the set $\text{pa}(Y_t(r))$. These conditional independence relationships at each time point t can be usefully represented pictorially by a directed acyclic graph (DAG) (see for example, Smith (1990)), where there is a directed arc to $Y_t(i)$ from each of its parents. We shall call any series which does not have any parents a *root node* and list all root nodes before any children in the ordered series $\boldsymbol{Y}_t$. For example, Figure 1 shows a DAG for five time series at time t, where $\text{pa}(Y_t(2)) = \varnothing$, $\text{pa}(Y_t(3)) = \{Y_t(1), Y_t(2)\}$, $\text{pa}(Y_t(4)) = \{Y_t(3)\}$ and $\text{pa}(Y_t(5)) = \{Y_t(3), Y_t(4)\}$. Both $Y_t(1)$ and $Y_t(2)$ are root nodes.

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

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

The vector $\boldsymbol{F}_t(r)$ contains the parents $\text{pa}(Y_t(r))$ and possibly other known variables (which may include $\boldsymbol{Y}^{t-1}(r)$ and $\text{pa}(\boldsymbol{Y}^{t-1}(r))$); $\boldsymbol{\theta}_t(r)$ 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 $\boldsymbol{w}_t^T = \{\boldsymbol{w}_t(1)^T, \ldots, \boldsymbol{w}_t(n)^T\}$, are such that $v_t(1), \ldots, v_t(n)$ and $\boldsymbol{w}_t(1), \ldots, \boldsymbol{w}_t(n)$ are mutually independent and $\{\boldsymbol{v}_t, \boldsymbol{w}_t\}_{t\ge1}$ 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 $Y_t(r)|\text{pa}(Y_t(r))$, $r = 2, \ldots, n$. 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. Forecasts for $Y_t(1)$ and $Y_t(r)|\text{pa}(Y_t(r))$, $r = 2, \ldots, n$, can also be found separately using established DLM results (see West and Harrison (1997) for details).

As the LMDM uses contemporaneous variables as regressors, we need the marginal forecasts for $Y_t(1), \ldots, Y_t(n)$. Unfortunately, the marginal forecast distributions will not generally be of a simple form. However, the marginal forecast means, variances and covariances are adequate for forecasting purposes and these can be calculated.

The forecast covariance matrix is not only of interest in its own right, but the marginal forecast covariance matrix for $\text{pa}(Y_t(r))$ is also required for calculating the marginal forecast variance for $Y_t(r)$. 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.

The next section presents a simple algebraic form for calculating the marginal forecast covariances.

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 $i < r$,*

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

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 forecast covariance matrix recursively.

Corollary 1 *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 forecast covariance using Theorem 1 and Corollary 1, 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. ■

It is possible for the DAG for an LMDM to contain deterministic nodes as well as random variable nodes (see, for example, Queen (1994) and Queen and Wright (2005)).

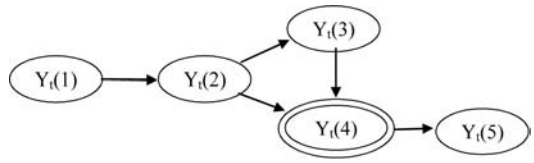

Figure 2: DAG with a deterministic node $Y_t(4)$: ◯ is a random variable, ◎ is a deterministic variable.

Figure 2 shows such a DAG where $Y_t(4)$ is a deterministic node. A deterministic node will be some (deterministic) function of its parents. If the function is linear, then the covariance between $Y_t(i)$ and a deterministic 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 2, suppose that $Y_t(4) = Y_t(3) - Y_t(2)$. Then

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

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

4 Component covariances

Suppose that $Y_t(r)$ has the m_r parents $Y_t(j_1), \ldots, Y_t(j_{m_r})$ for $r = 2, \ldots, n$. Write the observation equation for each $Y_t(r)$ as the sum of regression components as follows.

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

with

$$\begin{aligned} Y_t(r, j_l) &= Y_t(j_l)\theta_t(r)^{(j_l)} \\ Y_t(r, \boldsymbol{x}_t(r)) &= \boldsymbol{x}_t(r)^T \boldsymbol{\theta}_t(r)^{(\boldsymbol{x}_t(r))} \end{aligned}$$

where $\theta_t(r)^{(j_l)}$ is the parameter associated with the parent regressor $Y_t(j_l)$ 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 $Y_t(i,k)$ and $Y_t(r,j)$ from the models for $Y_t(i)$ and $Y_t(r)$ respectively. Call $\text{cov}(Y_t(i,k), Y_t(r,j))$ the *component covariance* for $Y_t(i,k)$ and $Y_t(r,j)$. We shall illustrate why component covariances may be useful by considering two examples.

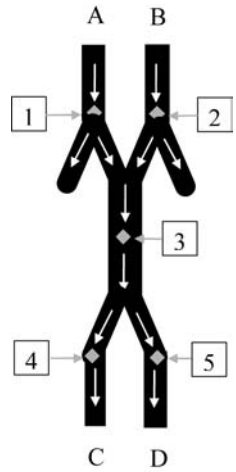

Figure 3: Hypothetical traffic network for Example 2 with starting points A and B and destinations C and D. The grey diamonds are the data collection sites, each of which is numbered. White arrows indicate the direction of traffic flow.

Example 2 Component covariances in traffic networks

Queen and Wright (2005) 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 3. 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 4. (For details on how this DAG can be elicited see Queen and Wright (2005).) Notice that all vehicles at site 3 flow to sites 4 and 5 so that conditional on $Y_t(3)$ and $Y_t(4)$, $Y_t(5)$ is deterministic with $Y_t(5) = Y_t(3) - Y_t(4)$.

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

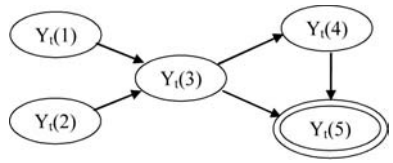

Figure 4: DAG representing the conditional independence relationships between $Y_t(1), \ldots, Y_t(5)$ in the traffic network of Figure 3: ◯ is a random variable, ◎ is a deterministic variable.

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 and Wright (2005) 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 5 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

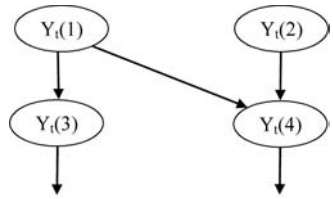

Figure 5: Part of a DAG for Example 3 for four time series $Y_t(1), \ldots, Y_t(4)$ (the DAG continues with children of $Y_t(3)$ and $Y_t(4)$).

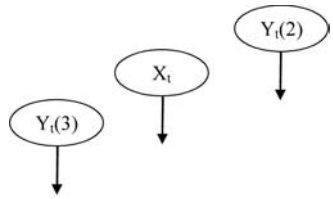

Figure 6: New DAG for Example 3 (adapting the DAG in Figure 5) where X_t is introduced into the DAG and $Y_t(1)$ and $Y_t(4)$ are no longer observed (the DAG continues with children of all three series).

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 5. Suppose further that $Y_t(1)$ and $Y_t(4)$ are no longer observed. The new DAG is given in Figure 6 (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 $Y_t(k)$ is the parent of $Y_t(i)$ and $Y_t(j)$ is a parent of $Y_t(r)$, with $i < r$. Then,*

$$cov(Y_t(i,k), Y_t(r,j)|D_{t-1}) = cov(Y_t(k), Y_t(j)|D_{t-1}) a_t(i)^{(k)} a_t(r)^{(j)}$$

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

Proof. For each r, let

$$\boldsymbol{Z}_t^{(r)^T} = \begin{pmatrix} Y_t(r,j_1) & Y_t(r,j_2) & \dots & Y_t(r,j_{m_r}) & Y_t(r,\boldsymbol{x}_t(r)) & v_t(r) \end{pmatrix}$$

and

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

Using the result that for two random variables X and Y, $\mathrm{E}(XY) = \mathrm{E}(X \cdot \mathrm{E}(Y|X))$, we have, for a specific $j \in \{j_1, \dots, j_{m_r}\}$,

$$\begin{aligned} \mathrm{cov}(\boldsymbol{Z}_t(r), Y_t(r,j)|D_{t-1}) &= \mathrm{E}(\boldsymbol{Z}_t(r) \cdot \mathrm{E}(Y_t(r,j)|\boldsymbol{Z}_t(r), D_{t-1})|D_{t-1}) \\ &\quad -\mathrm{E}(\boldsymbol{Z}_t(r)|D_{t-1})\mathrm{E}(Y_t(r,j)|D_{t-1}). \end{aligned} \tag{4.4}$$

Now, from Equation 4.1,

$$\mathrm{E}(Y_t(r,j)|\boldsymbol{Z}_t(r), D_{t-1}) = \mathrm{E}(Y_t(r,j)|Y_t(1), \dots, Y_t(r-1), D_{t-1}) = Y_t(j)a_t(r)^{(j)}.$$

So, using the result that for two random variables X and Y, $\mathrm{E}(Y) = \mathrm{E}(\mathrm{E}(Y|X))$, Equation 4.4 becomes,

$$\begin{aligned} \mathrm{cov}(\boldsymbol{Z}_t(r), Y_t(r,j)|D_{t-1}) &= \mathrm{E}(\boldsymbol{Z}_t(r)Y_t(j)a_t(r)^{(j)}|D_{t-1}) \\ &\quad -\mathrm{E}(\boldsymbol{Z}_t(r)|D_{t-1})\mathrm{E}(Y_t(j)a_t(r)^{(j)}|D_{t-1}) \\ &= \mathrm{cov}(\boldsymbol{Z}_t(r), Y_t(j)|D_{t-1})a_t(r)^{(j)}. \end{aligned}$$

Picking out the single row from $\boldsymbol{Z}_t(r)$ corresponding to $Y_t(i,k)$ gives us,

$$\mathrm{cov}(Y_t(i,k), Y_t(r,j)|D_{t-1}) = \mathrm{cov}(Y_t(i,k), Y_t(j)|D_{t-1})a_t(r)^{(j)}. \tag{4.5}$$

Let

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

Then

$$\begin{aligned} \mathrm{cov}(\boldsymbol{X}_t(r,i), Y_t(i,k)|D_{t-1}) &= \mathrm{E}(\boldsymbol{X}_t(r,i) \cdot \mathrm{E}(Y_t(i,k)|\boldsymbol{X}_t(r,i), D_{t-1})|D_{t-1}) \\ &\quad -\mathrm{E}(\boldsymbol{X}_t(r,i)|D_{t-1})\mathrm{E}(Y_t(i,k)|D_{t-1}). \end{aligned} \tag{4.6}$$

Now,

$$\mathrm{E}(Y_t(i,k)|\boldsymbol{X}_t(r,i), D_{t-1}) = Y_t(k)a_t(i)^{(k)},$$

so Equation 4.6 becomes

$$\text{cov}(\boldsymbol{X}_t(r,i), Y_t(i,k)|D_{t-1}) = \text{cov}(\boldsymbol{X}_t(r,i), Y_t(k)|D_{t-1})a_t(i)^{(k)}.$$

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

$$\text{cov}(Y_t(j), Y_t(i,k)|D_{t-1}) = \text{cov}(Y_t(j), Y_t(k)|D_{t-1})a_t(i)^{(k)}.$$

Thus Equation 4.5 becomes

$$\text{cov}(Y_t(i,k), Y_t(r,j)|D_{t-1}) = \text{cov}(Y_t(k), Y_t(j)|D_{t-1})a_t(i)^{(k)}a_t(r)^{(j)}$$

as required. ■

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

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

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

5 Covariance between root nodes

Recall from Corollary 1 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 7.

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,

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

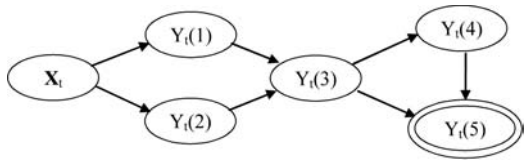

Figure 7: DAG representing time series of vehicle counts in Example 4, with an extra node $\boldsymbol{X}_t$ to account for any correlation between $Y_t(1)$ and $Y_t(2)$: ◯ is a random variable, ◎ is a deterministic variable.

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.

References

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

Queen, C.M. and Wright, B.J. (2005) Eliciting a DAG for a multivariate time series of vehicle counts in a traffic network. *Statistics Department Technical Report 05/07*, The Open University. (Available at http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm)

Smith, J.Q. (1990) Statistical principles on graphs. In *Influence Diagrams, Belief Nets and Decision Analysis*, R.M. Oliver and J.Q.Smith eds. John Wiley and Sons Ltd.

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